COMPUTER MODELLING AND
NEW TECHNOLOGIES

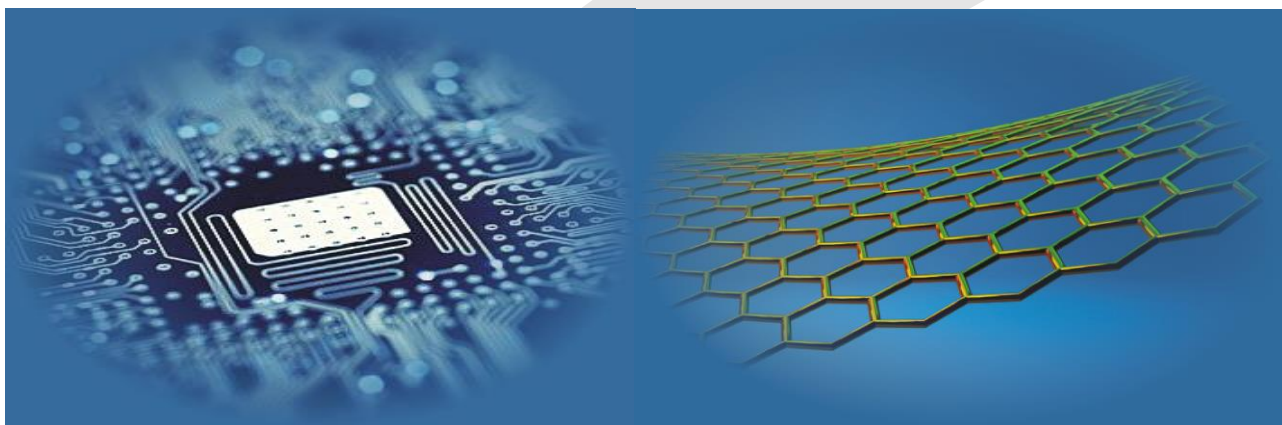# Computer Modelling and New Technologies

2014 Volume 18 No 9

Journal topics:	Publisher	Supporting Organizations
- mathematical and computer modelling - computer and information technologies - natural and engineering sciences - operation research and decision making - nanoscience and nanotechnologies - innovative education	Latvian Transport Development and Education Association	Latvian Academy of Sciences Latvian Operations Research Society Transport and Telecommunication Institute, Latvia Fraunhofer Institute for Factory Operation and Automation IFF, Germany International IT University, Kazakhstan

Articles should be submitted in **English**. All articles are reviewed.

Content

Nature Phenomena and Innovative Engineering

Authors' Index

Cumulative Index

Editors' Remarks

****

A British-Roman Song

by Rudyard Kipling

(A. D. 406)
"A Centurion of the Thirtieth"

My father's father saw it not,
And I, belike, shall never come
To look on that so-holy spot --
That very Rome --

Crowned by all Time, all Art, all Might,
The equal work of Gods and Man,
City beneath whose oldest height --
The Race began!

Soon to send forth again a brood,
Unshakable, we pray, that clings
To Rome's thrice-hammered hardihood --
In arduous things.

Strong heart with triple armour bound,
Beat strongly, for thy life-blood runs,
Age after Age, the Empire round --
In us thy Sons
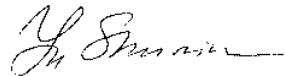
Who, distant from the Seven Hills,
Loving and serving much, require
Thee -- thee to guard 'gainst home-born ills
The Imperial Fire!

Rudyard Kipling (**1809-1849**) ♣

****
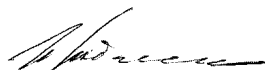
This 18th volume No.9 presents actual papers on main topics of Journal specialization, namely, **Mathematical and Computer Modelling, Computer and Information Technologies, Operation Research and Decision Making and Nature Phenomena and Innovative Engineering.**

Our journal policy is directed on the fundamental and applied sciences researches, which are the basement of a full-scale modelling in practice. This edition is the continuation of our publishing activities. We hope our journal will be interesting for research community, and we are open for collaboration both in research and publishing. We hope that journal's contributors will consider the collaboration with the Editorial Board as useful and constructive.

EDITORS

Yuri Shunin

Igor Kabashkin

♣ **Joseph Rudyard Kipling** (30 December 1865 – 18 January 1936) was an English short-story writer, poet, and novelist. He is chiefly remembered for his tales and poems of British soldiers in India and his tales for children. He was born in Bombay, in the Bombay Presidency of British India, and was taken by his family to England when he was five years old. Kipling is best known for his works of fiction, including The Jungle Book (a collection of stories, which includes and his poems, including "Mandalay" (1890), "Gunga Din" (1890), "The Gods of the Copybook Headings" (1919), "The White Man's Burden" (1899), and "If—" (1910). He is regarded as a major "innovator in the art of the short story"; his children's books are enduring classics of children's literature; and his best works are said to exhibit "a versatile and luminous narrative gift".

Distributed systems software architecture modelling and research methods

Xiaochun Zhao*

Heilongjiang University of Science and Technology

Abstract

With the development of computer network technology, the open, heterogeneous and distributed systems have become the mainstream in current computer applications because of the sharing resources, high availability, parallel processing and so on. However, due to the problems of development, which are constant expansion of systems size, evolution and continuous improvement, maintenance that required, specific distribution, autonomy and heterogeneity, a lot of research and software development practice shows, the introduction of software architecture which guide distributed system to develop and assume component blueprint is a practical and effective way to solve the difficulties of the development of distributed systems and build distributed systems successfully. Therefore, how to improve the quality and efficiency of distributed systems development by using software architecture, and ensuring system maintenance and space evolution are the key to develop distributed systems, also the core of this study. Software architecture, formal description of distributed systems interaction style, refinement and mapping architecture, distributed architecture systems development methods under evolution and reconstruction driving were studied based on the current distributed systems development methods as well as the problem of inadequate means.

Keywords: Software architecture, distributed systems, interaction style, software architecture refinement

1 Introduction

The main question in distributed system development is: the complexity of design, construction, commissioning, configuration and maintenance distributed applicant system is high in a heterogeneous environment, and causing high costs and low efficiency in developing, large-scale distributed systems development seems a risky challenge. With the component development ideas become mainstream in software development, people gradually realize that the software architecture is an important mean to control software complexity, improve the quality of software systems, support software development and reuse.[1] Therefore, it has immediate practical significance to in-depth research on the software system architecture, explore effective large-scale distributed systems development method, system design, analysis, and tectonic environment of architecture system driven, which help support its entire life cycle.

In this paper, a distributed system software architecture modelling and developing methods have been studied mainly from the following aspects: (1) propose a system architecture abstract model DSAM distributed which is suited system architecture , and give its formal model, on this basis, design and implement an attribute grammar-based software architecture description language DISADL and description language Discid based on CCS distributed component interaction style; (2) propose architecture refinement guiding principles and

design a set of mapping rules from a software architecture description DISADL to universal design model UML of software implementation; (3) propose a software architecture reconstruction based on fuzzy clustering analysis; (4) present distribution systems development method of the architecture driven, ADISC, establish its life cycle model, and put into system architecture supporting; (5) design and implement a visual supporting modelling environment for distributed software architecture structure, protocols, analysis, refinement, remodelling design, ADisDTool.

2 A Distributed software architecture description core model-DSAM

Model DSAM as a basis is designed a distributed software system architecture description language with component based by extending the traditional attribute grammars.

2.1 DSAM

Model DSAM includes: Event; Port; Interface; Connector Type; Component Type; Architecture Model. Figure 1 is a graphical representation of DSAM.

---

* *Corresponding author* e-mail: 986889936@qq.com
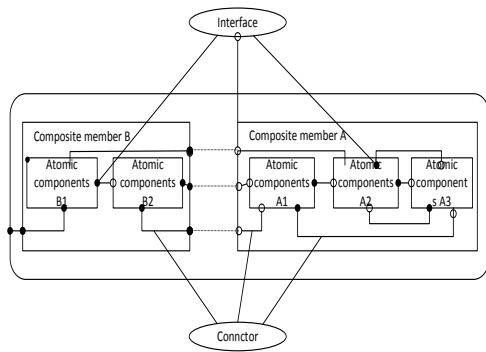
FIGURE 1 Architecture models DSAM

2.2 DESCRIPTION LANGUAGE DIS_ADL

2.2.1 Extending Attribute Grammar of Dis_ADL (EAG)

EAG is a seven-tuple: EAG = (T, N, P, Z, V, F, B), where T is the terminator collection; N is non-terminal symbol set; P is the production; Z is the starting character; V is the attribute range; F is calculation rules for the property; B is a finite set of conditions.

2.2.2 Semantic Description of Dis_ADL Based on EAG

Semantic description based on EAG includes: (a) logic timing and dependencies terminator; (2) parallel description terminator; (3) terminator @ (4) conditions production; (5) time constraints; (6) specific terminator

2.3 DESCRIPTION LANGUAGE - DISCID

Discid was meant to describe component interaction styles and verify formal nature. Discid begins with "Discid". Practical application shows, combination with Dis_ADL and Discid use various forms of mechanisms, describe the architecture from different viewpoints and form an organic whole, which greatly reduce the distributed system designer's cognitive difficulty, and also greatly improve the efficiency of the system design. Meanwhile, DIS_ADL and Discid are uncertainty and incompleteness in modelling, moreover, its adaptive capacity needs to be improved, Discid description needs further implementation language for mapping.

3 Models UML

UML is the most common object-oriented modelling language, modelling the software system by graph mode which from static structure and dynamic behaviour.

3.1 ARCHITECTURE MAPPING FOR

Distributed systems development with the guidance of software architecture build software systems architecture model by DIS_ADL, after the high-level completion of verification, seek to design fine from the top layer

constantly, up to a certain size, take the UML as the middle part the DIS_ADL architecture description is mapped to the appropriate elements in UML, and then developed into the design and implementation phase. Therefore, DIS_ADL and UML are combined and complement, DIS_ADL focus on high-level grasp and semantics and depth of precise, UML is based on a viable practice.

3.2 MAPPING RULES FROM DISADL TO UML

Conversion rules from Dis_ADL architecture description elements to UML elements are shown in Table 1.

TABLE 1 Mapping table of Dis_ADL element to UML elements

Element--Dis_ADL	Element--UML1.x	Element--UML2.0
Component Name	Class Name,Package Name	Class Name,Package Name
Atomic Components	Class Diagrams,Packages	Component,Packages,Class Diagrams
Composite member	Aggregation,composition, package, subsystem layer	Composite structures、Subsystem
State variables	Class private property	Class private property
Total Interface	Interface	Component supply Interface
Demand Interface	Abstract class	Component Requirements Interface
Port	Class method names, parameters, return type	Port
Message return type	Class method return type	Class method return type
Message parameters	Class method parameter (variable + type)	Class method parameter (variable + type)
Connector	Class diagrams, association	Connectors, class diagram, association
Architecture configuration	Component diagram or class diagram	Component diagram, structured category
Binding	Object Linking (associated instances)	Object Linking (associated instances)
Comment	Document	

4 Analysis of distributed software architecture reconstruction

4.1 DEFINITION OF SOFTWARE ARCHITECTURE RECONSTRUCTION [3]

Software architecture reconstruction is the process that reverses extract the architecture from being achieved or already implemented systems, reflecting the "actual construction" software architecture. The core problem is extraction of the architecture and the assessment the architecture evolution.

4.2 SOFTWARE ARCHITECTURE RECONSTRUCTION MODEL

Software architecture reconstruction process can be divided into four parts: (l) confirm stakeholders, defining target point of view, and the view collections. (2) extract

the underlying architecture element information from a variety of data sources (such as source code, documentation, etc.). (3) Define and process the extracted information, coordinate and establish connections between elements, generate a cohesive view of the architecture. (4) Construct data abstraction to generate architecture representation.

5 Distributed system design method -- ADISc and distributed development tectonic environment – AdisDTool

5.1 DISTRIBUTED SYSTEM DESIGN METHOD OF ARCHITECTURE DRIVEN – ADISC [4]

The development process of distributed systems can be divided with ADiSC into: requirements analysis; software architecture design, modelling and refinement conversion; system detailed design; system implementation, component assembly deployment and reconfiguration; system evolution and reuse.

5.2 DISTRIBUTED DEVELOPMENT TECTONIC SETTING --ADISDTOOL

Modelling system of ADisDTool consists of interactive graphics interpreter, DISADL language converter, DISADL lexer / parser, verify the nature of the software architecture, system builder, Discid language compiler environment, refinement converter, UML mapping generator, architecture reconstruction and other components, the system model is shown in Figure 2.
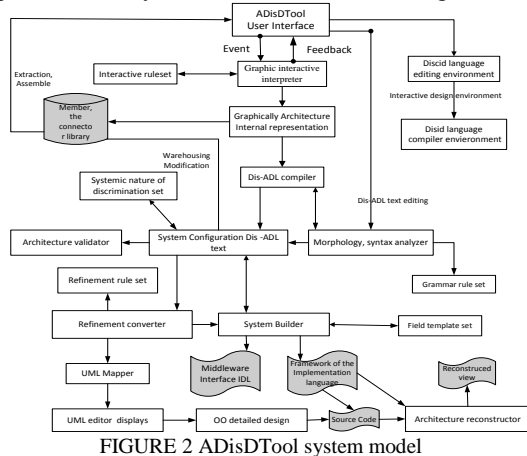
FIGURE 2 ADisDTool system model

5.3.1 Functional design of ADisDTool architecture modelling tool

The ADisDTool can be divided into eight modules: Project Management module, component and connector management module, visual modelling module architecture view modules, code generation module, system properties verification module, the system refinement mapping module, reconfigurable architecture.

5.3.2 ADisDTool reusable component library design [5]

ADisDTool established member (connector) library, which provided for the entire life-cycle management member. Provides the following functions: components, connectors, interfaces, warehousing; components, connectors, interfaces, query; components, connectors, export, evaluation, life cycle management, version control and so on.

6 Example authentication

Through such typical case of large-scale distributed systems development "digital content security platform based on DRM (National Innovation Fund project, project number 07c26226101995)", showing the core areas of applying ADISC methods to analyse, design and develop, providing convenient and effective design environment through ADisDTool, which does help convenient, fast and efficient analysis of large-scale distributed systems, nature verification, design, evolution and reuse, and prove effectiveness of DSAM models and ADISC method the thesis mentioned .

6.1 AS FOR "ORM-BASED DIGITAL CONTENT SECURITY PLATFORM."

The planform of Digital Content Protection Based on DRM is a system, which is analysed, designed, developed and achieved by distributed system, and the current one has been put to use for network environment of digital content security protection. PlatDRM ensured security of digital content in creating, distributing, using, sharing that throughout the life cycle. This paper briefly describes PlatDRM to validate the model DSAM and ADiSC method.

6.2 PLATDRM CORE FUNCTIONAL REQUIREMENTS

PlatDRM core functional requirements are: (1) to ensure digital content distribute secure, digital content is held being only in the specified environment and key user access, any duplicate is not available. (2) real-time authorization control, and can be changed the digital content access which have been distributed out at any time. (3) real-time monitoring and recording digital content action to take for the audit analysis, behaviour tracking provides detailed historical data. (4) support the parties authentication with domain authentication and other authentication systems integration, users only need to open an important document domain authentication.

6.3 TOP-LEVEL ARCHITECTURE DESIGN

According to ADiSC methods, requirement analysis selected core use cases; the resulting model is transformed to the software architecture to build a

language-independent architecture model. Meanwhile, it can be application-oriented modelling, describe domain business needs to facilitate greater reuse. PlatDRM is a typical distributed application, Figure 3 is its top-level conceptual architecture diagram, architecture consists of a client component, server component, WEB management component, of which the client is called CC member, the server is called SC members, managing client is called MC member.
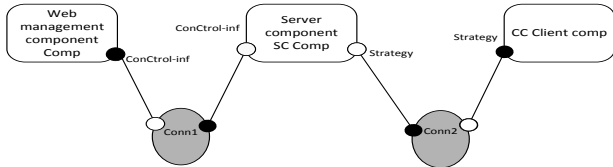
FIGURE 3 PlatDRM top Software Architecture Design

6.4 ARCHITECTURE REFINEMENT

When architecture is in a design stage, we use ADisDTool to analyse and refine PlatDRM architecture. First, for the interface refinement, the interface refinement between customer component and service component are set to be strategy interfaces, authentication interface, the operation log pick date, and digital content object interface. Then the client component CC internal structure refinement. After refining, Cc component architecture is shown in Figure 4.
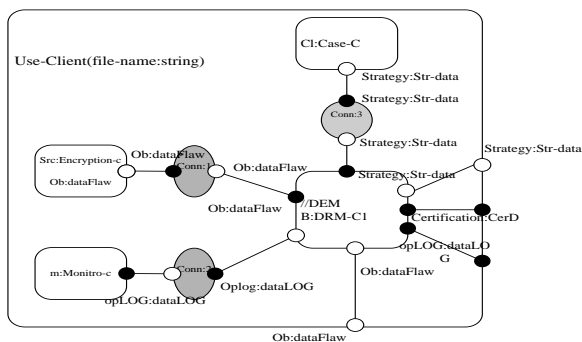
FIGURE 4 PlatFORM client component architecture after refinement

6.5 INTERACTION EVENTS CLIENT DEFINITION LOG_EVENT_CLIENT

After customers receiving log events, should be temporarily stored in the queue used to prepare the client component. LOG_EVENT_IO process defines the client endpoint behaviour, describes the state changing method after in response of API, also shows that when customers in the state of receiving start LOG_RECEIVED log events , the receiving message can only be placed in the event queue. To enable the server can send message directly at any time, avoiding the synchronization aborting between the server and the client, the definition of time endpoint contains two explicit intermediate state: STOP_LOG_EVENT_OUT   and STOP_LOG_EVENT_ACK, instead of the direct state transition from LOG_RECEIVED to LOG_NEGLECT.

6.6 MAPPING SYSTEM

In the system design phase, we designed classes, interfaces, components which PlatDRM containing by using UML. In addition, object state diagrams, sequence diagrams, etc. to get the attributes and methods of classes. Figure 5 is PlatDRM overall package design. Figure 6 is a design package diagram after client finer.
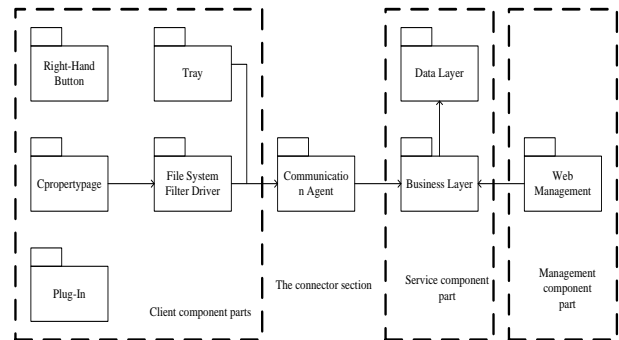
FIGURE 5 PlatDR, I overall design package diagram
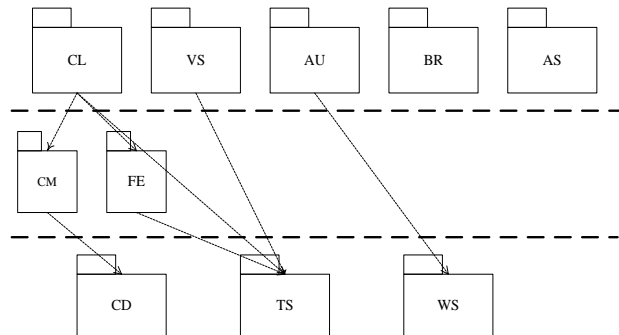
FIGURE 6 PlatDRM client design

6.7 PLATDRM FUZZY CLUSTERING ANALYSIS

Fuzzy clustering analysis, the results are shown in Figure 7.

		AS	AU	BR	CC	CD	CM	DP	DM	FE	RM	SU	TS	TU	WS	VS
													Client			
1	AS	Y														
2	AU		Y						Y							
3	BR			Y												
4	CC				Y			Y	Y		Y	Y		Y		
5	CD				Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y
6	CM				Y		Y	Y		Y						
7	DP				Y			Y								
8	DM				Y				Y							
9	FE				Y					Y						
10	RM							Y			Y					
11	SU							Y	Y		Y	Y				
12	TS				Y		Y	Y	Y	Y	Y	Y	Y	Y		Y
13	TU												Y	Y		
14	WS		Y												Y	
15	VS															Y

FIGURE 7 PlatDRM client component dependency matrix after refactoring

6.8 ANALYSIS AND RESULTS BY USING THE DSM CORRELATION ALGORITHM

Using the DSM-correlation algorithm, we can get matrix, shown in Figure 7, so we focused on {DP, DM, RM, SU, TU, CC} can be polymerized to obtain Figure 8. Relationship between the main clients' components can be clearly seen.

	BR	AU	WS	FE	DP	DM	RM	SU	TU	CC	CM	TS	VS	AS	CD
BR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AU	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
WS	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
FE	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
DP	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
DM	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
RM	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
R = SU	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
TU	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
CC	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0
CM	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
TS	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
VS	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
AS	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

FIGURE 7 PlatDRM client component reachable matrix view after refactoring

Client		BR	AU	WS	FE	CL	CM	TS	VS	AS	CD
1	BR	Y									
2	AU		Y			Y					
3	WS		Y	Y							
4	FE				Y	Y					
5	CL					Y					
11	CM				Y	Y	Y				
12	TS				Y	Y	Y	Y	Y		
13	VS								Y		
14	AS									Y	
15	CD				Y	Y	Y	Y	Y		Y

FIGURE 8 Client DSM after the polymerization

Analyse the results of DSM after client code division. After refreshing, we discovered that according to results of the matrix analysis, Figure 9 shows a PlatDRMclient's system architecture and Figure 8 is the corresponding matrix. Compared the figure represented with the original architectural design (Figure 6), it is easy to see the dependencies between components changed, and dependency occurred between CL and AU, contrary the initial design expectations, therefore, it needs to be adjusted and avoided.

FIGURE 9 P1atDRM client view after polymerization system packing

"Digital content security platform based on DRM" Such a typical distributed system has been put into the core link in ADISC, Dis_ADL describes the software architecture and is given top-level component description of the system, then refined it, and give a formal and intuitive description for the language Discid we mentioned focus on the typical event interaction style and conduct of the verification. Finally, the code that system implementation has been found in the software architecture, fuzzy clustering analysis and analyse it by using the proposed design structure matrix theories. Currently, the system by using ADISC methods for designing and developing was put into use in Shan xi military giant, which has been got a good evaluation of the user.

7 Conclusion

In the analysis of the lacking and problems in current large-scale distributed systems development methods and means of supporting, this paper presents a method based on component structure distributed systems under the support of the software driver. Its core is a distributed software architecture abstraction model--DSAM, a description language supported such architecture modelling, validation, refinement, evolution and reverse extraction over the DSAM--DIS_ADL and a component interaction analysis verification language based on CCS - -Discid. On this basis, this paper presents a visual support for distributed software architecture modelling, refinement and reconstruction of distributed system design environment--ADisDTool.

References

[1] Mei Hong, Jun-rong Shen 2006 Software Architecture Research *Journal of Software* **17**(06) 1257 - 75
[2] Shi-xian Zhang, Li-fu Wang, Fu-qing Yang 2000 Systems development based on COTS component *Computer Science* **27**(1) 6 - 8
[3] Lv-ai Sun, Mao-zhong Jin, etc 2002 Software Architecture Research *Journal of Software* **13**(07) 1225 - 37

[4] Ding-yi Fang 2001 Distributed software architecture research based on component interaction mechanism *Northwestern Polytechnical human science, Xi'an*
[5] Jian-jiang Zhu 2001 Software reuse research based software component *Nanjing University of Aeronautics and Astronautics*

Authors

Xiaochun Zhao, born on February 6, 1976, Jilin Province of China

Current position, grades: lecturer
University studies: Bachelor degree was earned in major of computer science and technology, Heilongjiang University of Science and Technology in 2007.
Scientific interest: computer application, web application, Web database technology, information processing

Three-dimensional deployment algorithm based on ideal fluid dynamics model for mobile sensor networks

Jiguang Chen*, Huanyan Qian

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China

Abstract

In the paper, a three-dimensional deployment algorithm based on ideal fluid model for sensor networks is proposed. On this basis, the proposed ideal fluid model is analysed, and the concept of flow field model is applied in deployment of wireless sensor networks. Sensor networks are abstracted as ideal fluid, with nodes as fluid micelles. In the deployment process, motion of nodes follows momentum conservation law of fluid micelle. Moreover, a simulation experiment is performed in this paper with the proposed deployment algorithm as the experimental subject. Coverage and uniformity are 2 indexes employed to evaluate the performance of the proposed algorithm. Shown by the simulation result, the three-dimensional deployment algorithm based on ideal fluid model for sensor networks leads to good deployment effect.

Keywords: three-dimensional deployment, mobile sensor networks, ideal fluid model

1 Introduction

Sensor network deployment refers to the process of adopting proper algorithm to deploy nodes within the target area, so as to conform to certain specified demand. Network deployment is a pre-condition for sensor network to play its role, as well as a premise for the normal operation of network [1-3]. It determines the detection effect of sensor network in the target area, which may further influence the service quality of the network. Node deployment of wireless sensor network is one of the kernel problems in the field. A good deployment method may significantly improve the perception quality of network, reducing resource consumption, and extending service life of network. Related scholars have performed plenty of studies in this connection. Normally, deployment of indoor sensor network nodes often puts to use regular deployment approach [4, 5] and most deployment in outdoor environment adopts random deployment approach [6-8].

Owing to diversified forms wireless sensor network nodes, deployment methods may be divided into three types:

1) Wireless sensor networks completely comprised by static nodes may be deployed manually.

2) As for wireless sensor networks comprised by static and mobile nodes, the mobility of mobile nodes [9] may be utilized to eliminate dead zones in sensor networks.

3) Wireless sensor networks comprised by mobile nodes may be deployed automatically [10, 11]. Owing to the mobility of nodes, deployment of mobile nodes is quite flexible [12].

It is easier to deploy nodes via releasing, spraying or other relevant methods. Yet, in unknown, complicated, and rigor environment, it is impossible to perform deployment manually. Nodes may be unable to be distributed in appropriate positions, leading to overlapping or dead zones, which wastes resources and affects network service quality. By contrast, deployment of mobile nodes perfectly solves the problem. In the algorithm, mobile nodes are designed to move following certain rules, and to elude obstacle accordingly. As for this, deployment in unknown and complicated environment becomes possible. Three-dimensional deployment of sensor network nodes based on ideal flow field model proposed in this paper belongs to this category. In the beginning, nodes may be places at accessible positions. On this basis, nodes will be automatically deployment according to prescribed motion rules, and to elude obstacles when applicable.

Literature [13-15] introduces the concept of potential field, i.e. a deployment method based on potential field. This method takes nodes as virtual particles influenced by virtual forces. One of the virtual forces creates repulsion between node and obstacle, as well as between node and node. As for this, nodes will expand rapidly based on the compact distribution in the beginning, so as to reach maximum coverage of network. Apart from such repulsive force, nodes are also influenced by viscous friction, which guarantees the network with a final static balance, i.e. all nodes will stop moving eventually. When the environment is changed, the network will be reconfigured accordingly to conform to the changed environment, and to reach static balance again.

This method leads to goods coverage rate. However, there is a premise - all nodes have to be put in the deployment environment in advance, i.e. the initial static deployment. What's more, in the initial deployment, the

---

* *Corresponding author* e-mail: jiguang_chen01@126.com

number of nodes shall be given. However, in unknown environment, it is impossible to figure out how many nodes are required so as to realize the deployment. Shortage of nodes reduces the coverage, while excessive nodes lead to redundancy, wasting energy. In this paper, the deployment method proposed by Literature [13-15] is improved. As for this, it is unnecessary to know the number of nodes before deployment, and the improved method is proved with high coverage effect.

Literature [16, 17] put forward a Virtual Force Algorithm (VFA) based on the assumption that nodes are movable, with the aim to improve the coverage of sensor nodes after initial deployment. For a sensor network with explicit number of nodes, virtual force algorithm tries to maximize network coverage rate. The algorithm employs attractive force and repulsive force to determine the path and rate of nodes' virtual motion. Cluster header is designed to calculate the destination position of nodes. Nodes will move accordingly to complete the task, if there is an effective destination position determined.

However, node position in the method is calculated by cluster header, which quite complicated in large-scale sensor networks, and is hereby not applicable. What we need to designed is deployment method applicable to large-scale networks with rigor environment. According to virtual force algorithm, a certain acting force is assumed between nodes, which promote the successful deployment of nodes.

Literature [18] adopts the basic control mechanism from virtual elastic network to deployment sensor networks. In every virtual elastic network, each node is considered as a particle, which moves regularly under the influence of virtual force. If a pair of nodes is too close, the elastic force between them will pull them apart. If the distance between them is too large, the elastic force will then push them closer. This is similar to the previously mentioned deployment method based on virtual force, but is still different in some aspects. The algorithm makes use of elastic force between nodes to expand nodes from the original compact and dense state to the whole area. Decisions are made directly between nodes, free from cluster headers. Literature [19] proposed a deployment algorithm based on ideal fluid model. Inspired by fluid dynamics, the paper proposed a new method to deploy mobile sensor network in unknown area. According to the physical rule of ideal fluid model, mobile sensor network is considered as a sort of fluid, while sensor nodes as micro fluid micelles. In the paper, fluid variables are projected into sensor networks, with equation solution method, initial condition and physical boundary condition given. Based on Literature [19], the method is improved and expanded to 3D deployment field, reaching good deployment effect.

The rest part of this paper is organized as follows: In Section 2, we analyse the ideal fluid dynamics model of three-dimensional deployment for mobile sensor networks. In Section 3, we solve the model mentioned in Section 2 and establish the ideal fluid dynamics algorithm

of deployment. In Section 4, we simulate the ideal fluid dynamics algorithm of deployment by using computer software and evaluate its performance. Finally, in Section 5, we reach the main conclusions.

The rest part of this paper is organized as follows: in Section 2, we analyse the ideal fluid dynamics model of three-dimensional deployment for mobile sensor networks. In Section 3, we solve the model mentioned in Section 2 and establish the ideal fluid dynamics algorithm of deployment. In Section 4, we simulate the ideal fluid dynamics algorithm of deployment by using computer software and evaluate its performance. Finally, in Section 5, we reach the main conclusions.

2 Analysis of ideal fluid control model

Nodes in sensor networks are independent individuals. There is no friction between nodes. In this paper, sensor networks are taken as ideal fluid. Ideal fluid omits dissipation, viscous transport, mass diffusion and heat conduction. Fluid model adopted in this paper defines fluid as a set comprised by flowing infinitesimal fluid micelles. Moreover, the paper is designed to study node motion. For this reason, conservation of momentum shall be considered. In other words, momentum equation followed by fluid micelles is applied in sensor nodes, so as to make nodes move following the rule, so as to realize sensor network deployment.

In this paper, motion of nodes in three-dimensional space Ω is researched. Moreover, the velocity along direction x, y and z is separately represented with u, v and w. Hereby, the three-dimensional expression of Euler equation may be described with Equation (1). It establishes the relation between the force on ideal fluid and the accelerated velocity of fluid motion. This, it is the foundation to study motions of ideal fluid, as well as an important equation in fluid dynamics.

$$\begin{cases} f_x - \dfrac{1}{\rho}\dfrac{\partial p}{\partial x} = \dfrac{du}{dt} \\[2mm] f_y - \dfrac{1}{\rho}\dfrac{\partial p}{\partial y} = \dfrac{dv}{dt} \\[2mm] f_z - \dfrac{1}{\rho}\dfrac{\partial p}{\partial z} = \dfrac{dw}{dt} \end{cases} . \qquad (1)$$

Equation (1) is an ideal fluid momentum equation, where, ρ is fluid density, d/dt is derivative of velocity on time, p refers to fluid pressure, f_x, f_y and f_z are component forces along directions of x, y and z; u, v and we are velocity components of an infinitesimal fluid element along directions of x, y and z. Expanding the derivative of velocity on time:

Chen Jiguang, Qian Huanyan

$$\begin{cases} \dfrac{du}{dt} = \dfrac{\partial u_i^t}{\partial t} + u_i^t \dfrac{\partial u_i^t}{\partial x} + v_i^t \dfrac{\partial u_i^t}{\partial y} + w_i^t \dfrac{\partial u_i^t}{\partial z} \\[2mm] \dfrac{dv}{dt} = \dfrac{\partial v_i^t}{\partial t} + u_i^t \dfrac{\partial v_i^t}{\partial x} + v_i^t \dfrac{\partial v_i^t}{\partial y} + w_i^t \dfrac{\partial v_i^t}{\partial z} \\[2mm] \dfrac{dw}{dt} = \dfrac{\partial w_i^t}{\partial t} + u_i^t \dfrac{\partial w_i^t}{\partial x} + v_i^t \dfrac{\partial w_i^t}{\partial y} + w_i^t \dfrac{\partial w_i^t}{\partial z} \end{cases} . \qquad (2)$$

Putting Equation (2) in Equation (1), (superscript and subscript are directly added here) more detailed Euler motion differential equation is shown as (3):

$$\begin{cases} \dfrac{\partial u_i^t}{\partial t} = -(u_i^t \dfrac{\partial u_i^t}{\partial x} + v_i^t \dfrac{\partial u_i^t}{\partial y} + w_i^t \dfrac{\partial u_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial x}) + f_{i,x}^t \\[2mm] \dfrac{\partial v_i^t}{\partial t} = -(u_i^t \dfrac{\partial v_i^t}{\partial x} + v_i^t \dfrac{\partial v_i^t}{\partial y} + w_i^t \dfrac{\partial v_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial y}) + f_{i,y}^t \\[2mm] \dfrac{\partial w_i^t}{\partial t} = -(u_i^t \dfrac{\partial w_i^t}{\partial x} + v_i^t \dfrac{\partial w_i^t}{\partial y} + w_i^t \dfrac{\partial w_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial z}) + f_{i,z}^t \end{cases} . \qquad (3)$$

In Equation (3), time derivative is on the left, and space derivate is on the right. Subscript i and Superscript t represents fluid element i's parameters of u, v, w, p, ρ and f at the time of t.

3 Applying fluid concept to sensor networks

In order to make use of the liquidity of ideal fluid to analyse the deployment process of sensor network, the network model is assumed as follows:

1) Nodes shall be able to move: Nodes in the network shall be able to move freely, while the energy of nodes shall be able to support nodes to move for a long distance so as to realize the deployment, as well as affairs after the deployment. This is the most fundamental ability of nodes, as well as the basis for self-deployment of network.

2) Nodes shall be designed with the function of self-positioning. Nodes shall be able to know their current position and velocity in every phase. Such positioning function is necessary for each node to determine the next action in the deployment process.

3) Nodes shall be designed with the function to perceive the environment. The perception coverage with each sensor node may be considered as a circle with radius of R_s. As for this, nodes are able to perceive the position of obstacle, etc.

4) Nodes shall be designed with communication function within finite range. The communication coverage of each sensor node is defined by another circle with radius of R_c. A sensor node shall be able to perceive the relative position and velocity of its neighbour nodes within the communication range.

5) Nodes shall be designed with calculation and information fusion function. Useful information may be extracted for decision-making, controlling motion direction and velocity, so as to move freely.

Definition of fluid variables used in Equation (1)-(3) in wireless sensor networks is described in Table 1.

TABLE 1 Fluid Property and WSNs Property

Fluid Property	Symbol	WSNs Property
Support Domain	Ω	Local neighbourhood within communication range
Position	x	Position vector
Velocity	v	Velocity control vector
Density	ρ	A measure of the number of nodes in Ω
Pressure	p	A parameter generating repulsion among nodes
Body force	f	Flow control force vector

3.1 SOLUTION OF MODEL CONTROL EQUATION

Similar to computations in fluid dynamics, the velocity of sensor nodes may as well be figured out by time iteration. Taking Velocity Component u of Node i as the example, assuming that the value of velocity component at Time t is given, after a small time interval Δt, the value of velocity component may be calculated with Taylor expansion, so as to obtain the following equation:

$$u_i^{t+\Delta t} = u_i^t + (\dfrac{\partial u_i^t}{\partial t})\Delta t + (\dfrac{\partial^2 u_i^t}{\partial t^2})\dfrac{(\Delta t)^2}{2} + ... \qquad (4)$$

Velocity component may be simply and approximately described as the first two terms in Equation (4), i.e.:

$$u_i^{t+\Delta t} = u_i^t + (\dfrac{\partial u_i^t}{\partial t})\Delta t . \qquad (5)$$

Putting Equation (3) in Equation (5), after a small time interval Δt, the value of velocity component may be described as:

$$\begin{cases} u_i^{t+\Delta t} = u_i^t + (-(u_i^t \dfrac{\partial u_i^t}{\partial x} + v_i^t \dfrac{\partial u_i^t}{\partial y} + w_i^t \dfrac{\partial u_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial x}) + f_{i,x}^t)\Delta t \\[2mm] v_i^{t+\Delta t} = v_i^t + (-(u_i^t \dfrac{\partial v_i^t}{\partial x} + v_i^t \dfrac{\partial v_i^t}{\partial y} + w_i^t \dfrac{\partial v_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial y}) + f_{i,y}^t)\Delta t \\[2mm] w_i^{t+\Delta t} = w_i^t + (-(u_i^t \dfrac{\partial w_i^t}{\partial x} + v_i^t \dfrac{\partial w_i^t}{\partial y} + w_i^t \dfrac{\partial w_i^t}{\partial z} + \dfrac{1}{\rho_i^t}\dfrac{\partial p_i^t}{\partial z}) + f_{i,z}^t)\Delta t \end{cases} . (6)$$

In calculation, equations of u_i^t, v_i^t, w_i^t, ρ_i^t, and p_i^t may be put in Equation (6), so as to calculate the value of velocity component by iteration, and to further figure out the next position of Node i. In calculation, momentum conservation law is followed. Motion of nodes abides by the physical rule of fluid micelle.

3.2 CONSTRAINT CONDITION, INITIAL CONDITION AND PHYSICAL BOUNDARY CONDITION

The above has described node motion. However, when the motion begins, there shall be initial conditions give. Moreover, in the motion process, nodes' motion shall be

14

Chen Jiguang, Qian Huanyan

limited. When nodes reach the boundary or encounter obstacles, motion will also be changed. In the following, these problems are to be solved.

1) Initial condition.

Time iteration method in Equation (5) needs to know the initial value of velocity component in advance, as well as present position before calculating the next position. Before deployment, initial velocity and position of any node may be known.

2) Physical boundary condition.

For ideal fluid, as there is no friction, flow velocity on object plane is a finite non-zero value. Moreover, for impervious walls, there is no mass inflow or out-flow object plane. As for this, the velocity of fluid sticking close to object plane shall be tangent with object plane. In other words, the velocity component perpendicular to object plane shall be zero, i.e. flowing on object plane is tangent with object plane. This is the only object plane boundary condition of frictionless flow. The value of velocity on object plane, along with temperature, pressure and density of fluid on object plane, will become a part of the solution. As is shown by Figure 1, when the distance from node to obstacle or boundary is smaller than d, the velocity will be changed.
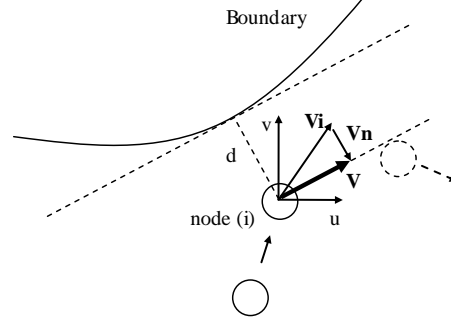
FIGURE 1 Object plane boundary condition

3) Constraint control condition.

Velocity component method in Equation (6) may be used to calculate the velocity of sensor node. In reality, owing to mobile equipment or other problems, the motion velocity of network sensor nodes may be limited. As for this, calculated velocity and accelerated velocity of sensor nodes shall be mandatorily constrained, in case that they exceed the threshold of velocity V_{th} and accelerated velocity a_{th}.

3.3 IDEAL FLUID DEPLOYMENT ALGORITHM

The entire Deployment Algorithm of Ideal Fluid Model is shown in Algorithm 1.

ALGORITHM 1 Deployment Algorithm of Ideal Fluid Dynamics

Algorithm : Deployment Algorithm of Ideal Fluid Model
Requirement: Ω, N, C, Rc, Rs.

1:	Estimate number of nodes N based on expected Coverage Ratio C and volume of Ω
2:	At time t_0, all of the N nodes are located at their initial positions
3:	For each node node_i at time t
4:	Parameters in Euler equation may be figured out by calculating the distance with neighbor nodes;
5:	Accelerated velocity a_i of node along different direction may be calculated with parameters obtained from 4;
6:	Calculating node velocity v_i;
7:	Judging if the velocity is larger than the prescribed threshold v_{th}; if so, turn to 8; if not, turn to 9;
8:	Changing node velocity to threshold velocity;
9:	Calculating the position of the next step;
10:	Judging if the distance from node to obstacle or boundary is smaller than the prescribed distance; if so turn to 11; if not turn to 12;
11:	Changing velocity according to boundary condition, and turning to 9;
12:	Node moves to the calculated position;
13:	Judging if the target coverage is reached; if so, turn to 14; if not, turn to 4;
14:	End For
15:	All nodes have reached equilibrium, algorithm ends.

With a 3D plot, the principle of the algorithm in this paper can be clearly demonstrated, as is shown in Figure 2. This is a 3D plot simulated with software, and the deployment area Ω is a space with the dimension of 10x10x10. The initial deployment position of node is a corner in the area, as is shown in the left plot. The small black dots are exactly the mobile nodes to be deployed. The right plot shows the spatial distribution of nodes after deployment. It can be seen from the plot that, nodes have been expanded from the initial positions to the whole space.
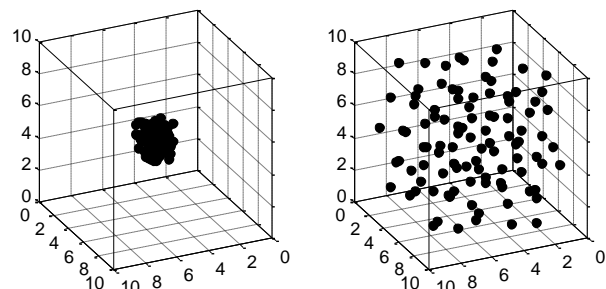
FIGURE 2 Demonstration of the Deployment Process

4 Simulation results

Numerous simulations have been carried out in various environments to investigate the performance of our approach. Parameter settings of our simulations are indicated in Table 2 with reference to the related figure number.

TABLE 2 Simulation settings by figure number

Parameter	Figure 3	Figure 4	Figure 5
Deployment(Ω)	$10\times10\times10$	$10\times10\times10$	$10\times10\times10$
Node number (NN)	100	100	100
R_s, R_c	2; 4	2; 4	2; 4
damping factor (Df)	2.0	2.0	2.0
a_{th}, v_{th}	2; 1	2; 1	2; 1
Time (T)	15	15	15
Δt	0.1	0.1	0.1
Obstacles	0	1	0
ROIs	0	0	1

4.1 COVERAGE AND UNIFORMITY

Generally, coverage can be considered as the service quality of a sensor network. Gage invented the concept of coverage in the research of multi-robot systems [20]. This paper defines it as the ratio between the sum of the coverage volume of all nodes and the volume of the entire target region, as is shown in Equation (7). The definition of the sum of the coverage volume is taken from the concept of union in the Set Theory, thus the coverage is usually no larger than 1.

$$Coverage = \frac{\bigcup\limits_{i=1,...,N}\Omega_i}{\Omega} . \tag{7}$$

The uniformity of coverage is a well-defined standard to measure the service life of a network. Article [21] describes the concept as the standard deviation of distance between nodes. Smaller standard deviation means better coverage uniformity of the network. However, this approach for measuring uniformity is under perfection. We take the grid approach to calculate uniformity in this paper, i.e. dividing the whole region into N small cubes with the same volume, and then figuring out the standard deviation of the nodes contained in these small cubes, as is shown in Equation (8).

$$U = \left[\frac{1}{N}\sum_{i=N}^{N}(n_i - \bar{n})^2\right]^{\frac{1}{2}} . \tag{8}$$

In the equation, n_i is the total number of nodes in the ith small cube and $\bar{n}$ is the mean of nodes number in each cube. So far, we have discussed the relation between communication and coverage. Article [22] has proved that when the communication range of node is twice or larger than the sensing range, coverage will contain pure connections. In practical deployment, we only have to

consider the coverage so as to ensure the connection. At the moment, coverage contains connection problems.

Please refer to Table 2 for simulation parameters in Figure 3. The circumstance after deployment is shown in the left plot. In the plot, the cube represents the volume of the region (area Ω). Tiny black dots represent the positions of nodes; blue spheres are used to indicate the sensing radius. In order to watch them clearly, only the sensing radius of partial nodes are displayed. Nodes can probe the environment and collect information in its sensing radius R_s. Similarly, it may be influenced by the Coulomb's force from its adjacent nodes, and repulsive force from the obstacles as well. The communication radius of Node R_c is greater than the sensing radius R_s. Nodes are able to exchange information mutually within the communication radius. The initial positions of the nodes are in the centre of Region Ω (being close to the coordinates [5,5,5]). Imaginably, node density at the beginning is practically very large. When deployment is started, all the nodes are forced by un-balanced Coulomb's force, starting to move to fill the entire region, and eventually reaching equilibrium, as is shown in Figure 3. The right plot shows the change in Coverage and Uniformity versus time during the deployment process described in the left plot. The x-coordinate is the simulation time T; the y-coordinate on the left is the uniformity; the y-coordinate on the right represents the coverage. We can see from the right plot that, at the initial moment of deployment, the nodes only cover a small part of the region because they are all initially positioned in the centre of Ω. Therefore the coverage is practically low (<20%). With time elapsed and all nodes are in motion, the coverage tends to grow, and reaches equilibrium after a certain moment (about Time=9). This begins when the coverage reaches maximum (about 100%), and lasts until the simulation is over. The value of uniformity is initially around 1.8 when the simulation begins. Then it decreases rapidly shortly after the deployment begins, eventually, tends to be around 0.3. Because uniformity represents standard deviations, so the smaller it is, the more uniform the network is.
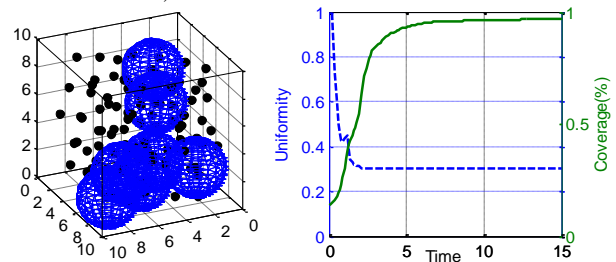
FIGURE 3 Normal deployment

In the left plot in Figure 4, the solid spheres represent obstacles. In the previous sections we mentioned that obstacle can be considered stationary charge with same sign. But the charge it carries is proportional to its size, which means the larger it is, the more charge it carries. There is only one obstacle in the left plot. We can adjust the distance from nodes to the obstacles by configuring the amount of charges that the obstacle carries. It is shown in

Chen Jiguang, Qian Huanyan

the left plot that, due to the repulsive force from the obstacles, nodes will be deployed at a distance with the obstacle to avoid them. This is highly meaningful in applications. To keep the nodes away from hazard or unreachable region can minimize node damage, indicating the algorithm's self-adaption. The right plot shows the change in coverage and uniformity versus time during the deployment process. We can see that because of the obstacles, the region is not completely covered by nodes. The maximum coverage is less than 1. In the meantime, compared with Figure 3, the value of uniformity is higher after nodes have reached equilibrium, due to reduction of uniformity because of obstacles.
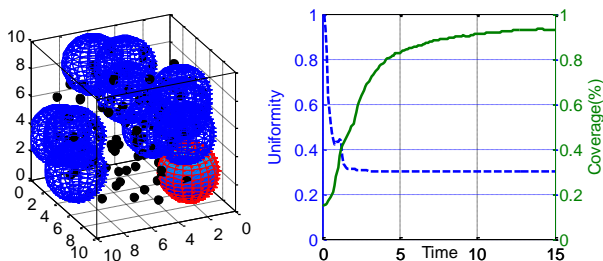
FIGURE 4 Deployment with Obstacles

In the left plot in Figure 5, the black little solid spheres represent sensor nodes while the three big spheres represent the ROIs (Range of Interesting). We noted that ROI can be considered stationary charge with opposite sign. While the charge ROI carries is proportional to its size, which means the larger it is, the more charge it carries. There are three ROIs in the left plot. We can adjust the distance from nodes to the ROIs by configuring the amount of charges that the ROI carries. It is shown in the left plot that, due to the attractive force from the ROIs, nodes will be deployed close to them. This is highly meaningful in applications. To put more attentions to the region needed monitor. The right plot shows the change in coverage and uniformity versus time during the deployment process. We can see that because of the ROIs, the region is completely covered by nodes. When the maximum coverage is reached, it decreases slowly with time for nodes tend to move toward to the ROIs. In the

meantime, compared with Figured 3, the value of uniformity is higher after nodes have reached equilibrium, due to reduction of uniformity because of ROIs.

FIGURE 5 Deployment with ROIs

5 Conclusion

Most previous researches on sensor networks are limited to 2d planes, neglecting the vertical fall of the deployment area, while taking the area as a flat plane. With people's continual exploration on space, underwater, underground and other regions, 3D sensor networks are gradually becoming a research hot spot. Yet, 2D algorithms are no longer applicable to 3D deployment. As for this, deployment algorithms under 3D environment have to be researched.

The proposed ideal fluid model is analysed in this paper, and the concept of flow field model is applied in deployment of wireless sensor networks. Sensor networks are abstracted as ideal fluid. In the deployment process, motion of nodes follows momentum conservation law of fluid micelle. A simulation experiment is performed in this paper with the proposed deployment algorithm as the experimental subject. Coverage and uniformity are 2 indexes employed to evaluate the performance of the proposed algorithm. Shown by the simulation result, the deployment algorithm based on ideal fluid model for sensor networks leads to good effect.

Acknowledgments

References

[1] Liu B, Dousse O, Nain P, Towsley D 2013 *Parallel and Distributed Systems IEEE Transactions on* **24**(2) 301-11
[2] Luo R C Chen O 2012 *Industrial Electronics IEEE Transactions on* **59**(5) 2377-85
[3] Xu K, Hassanein H, Takahara G, Wang Q 2010 *Mobile Computing, IEEE Transactions on* **9**(2) 145-59
[4] Chipara O, Hackmann G, Lu C, Smart W D 2010 Practical modeling and prediction of radio coverage of indoor sensor networks *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks ACM* 339-49
[5] Huchard M, Paquier V, Loeillet A 2012 Indoor deployment of a wireless sensor network for inventory and localization of mobile assets[C]//RFID-Technologies and Applications (RFID-TA) International Conference on IEEE 369-72

[6] Ozturk C, Karaboga D, Gorkemli B 2011 Probabilistic dynamic deployment of wireless sensor networks by artificial bee colony algorithm *Sensors* **11**(6) 6056-65
[7] Kulkarni R V, Forster A, Venayagamoorthy G K 2011 *Communications Surveys & Tutorials IEEE* **13**(1) 68-96
[8] Liao W H, Kao Y, Li Y S 2011 A sensor deployment approach using glowworm swarm optimization algorithm in wireless sensor networks *Expert Systems with Applications* **38**(10) 12180-8
[9] Wang G, Cao G, Berman P, La Porta T F 2007 *Mobile Computing, IEEE Transactions on* **6**(5) 563-76
[10] Yang F, Zhang Y A 2013 Distributed Deployment Algorithm with Mobile Sensors in Wireless Sensor Network *Advanced Technology in Teaching Springer Berlin Heidelberg* 533-40
[11] Zhang Y, Wang L A 2010 distributed sensor deployment algorithm of mobile sensor network *Intelligent Control and Automation (WCICA) 8th World Congress on IEEE* 6963-8

[12] Wang H, Guo Y 2008 A decentralized control for mobile sensor network effective coverage *Intelligent Control and Automation, WCICA 7th World Congress on IEEE* 473-8

[13] Howard A, Matarić M J, Sukhatme G S 2002 Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem *Distributed autonomous robotic systems 5 Springer Japan* 299-308

[14] Popa D O, Stephanou H E, Helm C, Sanderson A C 2004 Robotic deployment of sensor networks using potential fields *Robotics and Automation Proceedings ICRA'04 IEEE International Conference on IEEE* **1** 642-7

[15] Lee J, Dharne A D, Jayasuriya S 2007 Potential field based hierarchical structure for mobile sensor network deployment *American Control Conference ACC'07 IEEE* 5946-51

[16] Zou Y, Chakrabarty K 2003 Sensor deployment and target localization based on virtual forces *INFOCOM Twenty-Second Annual Joint Conference of the IEEE Computer and Communications IEEE Societies IEEE* **2** 1293-303

[17] Wang G, Cao G, La Porta T 2006 *Mobile Computing IEEE Transactions on* **5**(6) 640-52

[18] Shucker B, Bennett J K 2007 Scalable control of distributed robotic macrosensors *Distributed Autonomous Robotic Systems 6 Springer Japan* 379-88

[19] Pac M R, Erkmen A M, Erkmen I 2006 Towards fluent sensor networks: A scalable and robust self-deployment approach *Adaptive Hardware and Systems, AHS 2006 First NASA/ESA Conference on IEEE* 365-72

[20] Gage, Douglas W Command control for many-robot systems. *Naval command control and ocean surveillance center rdt and e div San Diego Ca* 1992

[21] Heo, Nojeong, Varshney K P 2003 A distributed self-spreading algorithm for mobile wireless sensor networks *Wireless Communications and Networking WCNC 2003 IEEE*. **3**.

[22] Alam, S M, Haas Z J 2006 Coverage and connectivity in three-dimensional networks *Proceedings of the 12th annual international conference on Mobile computing and networking. ACM*

Authors

Chen Jiguang, born in February, 1982, Henan Province, China

Current position, grades: Ph.D. candidate in school of computer science and engineering, Nanjing University of Science and Technology.
University studies: B.S. and M.S. degrees in computer science and technology from Henan Normal University, Xinxiang, Henan, China, in 2004 and 2008.
Scientific interest: sensor deployment and localization in mobile sensor networks and wireless sensor networks.

Qian Huanyan, born in October, 1950, Jiangsu Province, China

Current position, grades: Professor and PhD supervisor in school of Computer Science and Engineering, Nanjing University of Science and Technology.
University studies: B.S. degree in Computational mathematics from Nanjing University in 1977.
Scientific interest: sensor networks, mobile communication, wireless communication Networks.

Goal's three-dimensional trajectory reconstruction based on the adaptive multiple target surface iteration method

Ming Lei*, Mingming Chen, Hanshan Li

School of Electronic Information Engineering, Xi'an Technological University, Xi'an, Shaanxi 710021, China

Abstract

In order to solving the problem of the multiple bullet target matching and multiple bullet three-dimensional trajectory calculation., this article proposes a method for bullet three-dimensional trajectory measurement based on an adaptive multiple target surface and multiple iteration algorithm. In this method, it set up a virtual cuboids target space, which includes multiple bullet trajectories. It used two high speed cameras, one in left and another in right, to capture images. In the vertical direction of the bullet flying, these images were segmented into infinite plane, which are multiple target surfaces. It used the projective transformation to recover the target two-dimensional image from the left and the right images, and repeated this process until finding the bullet point in left view is coincide with that in right view. If the points were found, it indicated that it achieves the goal's accurately matching and every space location of the bullet. The simulation experimental results show that this method is feasible, image processing and analysis is merely influenced by background. This method can effectively realize the bullet three-dimensional trajectory target matching, and it has strong manoeuvrability.

Keywords: three-dimensional trajectory, target matching, perspective transformation, multiple iterations, adaptive multiple target surface

1 Introduction

In the field of the three-dimensional coordinates and the three-dimensional trajectory, a key question is how to achieve multi-target bullet points of the match of the two-dimensional image on the left and right cameras. How to automatically and efficiently improve the accuracy of matching that is an important link in achieving the bullet matching. The commonly used image matching technology is mainly divided into the method based on pixel and the method based on feature [1-3]. Based on the direct method is to directly calculate the difference value of image pixel grey value and that the same part of the pixel grey value should be inferior to different parts of the pixel grey value of value. This method is simple and intuitive, but the drawback is a large amount of calculation, and as the image has a larger under the condition of different rotation and illumination, the method cannot obtain satisfactory results. In order to reduce the grey-scale image matching method based on the false match rate and improve the matching algorithm noise is less affected by noise, so the method based on characteristic is presented. Based on the characteristics of the method is to extract contains important image feature points from the necessary matching image, and then use similarity measure for image feature points matching. The commonly used features are edge, outline linear and angular point in image matching. Matching based on feature for image distortion has certain robustness, but its matching performance largely depends on the quality of the image feature extraction, and this kind of method of image feature extraction are susceptible to

noise interference and influence [4, 5]. Due to the complexity of imaging in the scene condition as well as a variety of different sensor has the characteristics of different imaging mechanism, the features of the extraction of image stabilization point will sometimes be very difficult. But when handling the bullet target matching in the actual situation, if the background of image is relatively complex, so the extraction of feature points is inevitably affected by the noise and caused error rate in the extraction of feature points, which lead to the decrease of matching precision. So the above methods cannot fully applicable to the situation, it need to find a new method that is less affected by the environment, and can be accurate and efficient implementation of the target matching. The adaptive multiple target surface represents a planar target surface can be unlimited extension and the distance between the planes can be equidistance change. Thus, this paper proposes a new adaptive multiple target surface iteration method to achieve the objectives matching. By finding about camera imaging of left and right view the same target at the same time in the same plane to achieve the goal of matching. The method by the left and right view image perspective transformation [6-8] front view image then compare the pixels in the graph in the face to judge the pixels are consistent about the corresponding view of whether the bullet is the same goal. This method is rapid, easy to solve the matching problem of the bullet, thus calculated to the three-dimensional coordinate in the process of bullet flight.

Corresponding author e-mail: maxleiming@163.com

2 Adaptive target face multiple iterative method principle

2.1 ADAPTIVE MULTIPLE TARGET SURFACE MODEL

In reality, the bullet's flying process can be seen as the bullet through countless planes, which are parallel to each other. As shown in Figure 1, the bullet's flying state can be expressed as a process that the bullet flies through the first plane to the *N*-th planar. The plane is parallel to each other and the relative distance is very tiny. So three-dimensional coordinate received in each plane when the bullet flies through each plane can be used to represent the bullet's flying 3D trajectory figure.

FIGURE 1 The process of the bullet flying

Through the theoretical analysis above, we can adopt the virtual plane, which is made of 8 landmark point when the bullet flies through it. The plane *ABCD* is parallel to the plane *EFGH*, and the line *AD*, *BC*, *EH*, *FG* are horizontal and have equal spaces. If we divide the planes between the plane *ABCD* and plane *EFGH* in average, we can get *N* parallels planes and form random number of parallel planes. The bullet's flying through the construction plane is shown in Figure 2.

FIGURE 2 The bullet's flying through the construction plane

2.2 THE PRINCIPLE OF MODEL ANALYSIS

When we use two array high-speed cameras, which are placed in both the left and right sides to capture images of the bullet flight, the imaging process of each image can be seen stationary. Because of rotation and translation, orthography is formed on the image plane when we move the observation point and the image plane on condition that the object plane is stationary. The imaging process can be seen as a plane, which the actual orthography plane is mapped to at any angle. That is to say, the imaging process is the projective transformation, which transfers from a two-dimensional planar scene to another, the point, which transfers from one plane to another [9].

Therefore, the transformation can be expressed as *H*, which is a 3×3 holography regardless of global scale factor in the projective space [10-13]. Thus, *H* has 8 DOF. As shown in Equation (1), *M* is the homogeneous coordinates in the plane of the scene; m is the imaging coordinates on the image plane.

$$M = Hm. \tag{1}$$

Thus, according to the camera model, Equation (1) can be rewritten as Equation (2) by transformation. The coordinates on the plane of random point *M* is (*X*, *Y*, *Z*) the corresponding coordinates of image point is (*u*, *v*, 1). Through calculating multiple points, we can obtain holography transferring Equation (1) to Equation (2).

$$\begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -X_1u_1 & -X_1v_1 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -Y_1u_1 & -Y_1v_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & -X_2u_2 & -X_2v_2 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -Y_2u_2 & -Y_2v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n & v_n & 1 & 0 & 0 & 0 & -X_nu_n & -X_nv_n \\ 0 & 0 & 0 & u_n & v_n & 1 & -Y_nu_n & -Y_nv_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ \vdots \\ \vdots \\ X_n \\ Y_n \end{bmatrix}. \tag{2}$$

Assuming that the cameras on both sides film the same target, and adopting the unified reference coordinate system, then we can get that the coordinates (*X*, *Y*) of random point *M* on the scene plane are the same, while the imaging coordinates (*u*, *v*) on the imaging plane resulted from left and right view are different. Thus, the relationship between pixels of front view images and pixels of different angle images can be obtained from the deformation of Equation (4) and expressed as Equation (5). The coordinates (*u*, *v*) is on the imaging plane after the perspective transformation, while (*u'*,*v'*) is on the front view coordinates.

$$\begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -X_1u_1 & -X_1v_1 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -Y_1u_1 & -Y_1v_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & -X_2u_2 & -X_2v_2 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -Y_2u_2 & -Y_2v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n & v_n & 1 & 0 & 0 & 0 & -X_nu_n & -X_nv_n \\ 0 & 0 & 0 & u_n & v_n & 1 & -Y_nu_n & -Y_nv_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = $$

(3)

$$\begin{bmatrix} u_1' & v_1' & 1 & 0 & 0 & 0 & -X_1u_1' & -X_1v_1' \\ 0 & 0 & 0 & u_1' & v_1' & 1 & -Y_1u_1' & -Y_1v_1' \\ u_2' & v_2' & 1 & 0 & 0 & 0 & -X_2u_2' & -X_2v_2' \\ 0 & 0 & 0 & u_2' & v_2' & 1 & -Y_2u_2' & -Y_2v_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n' & v_n' & 1 & 0 & 0 & 0 & -X_nu_n' & -X_nv_n' \\ 0 & 0 & 0 & u_n' & v_n' & 1 & -Y_nu_n' & -Y_nv_n' \end{bmatrix} \begin{bmatrix} a1 \\ b1 \\ c1 \\ d1 \\ e1 \\ f1 \\ g1 \\ h1 \end{bmatrix}.$$

Thus, perspective transformation can be simplified as the relationship between pixels of front view images and image point of the plane image after the perspective transformation. As shown in Equation (3), A is pixel coordinates of image after perspective transformation. B is pixel coordinates of the front view, is the perspective transformation function.

$$A\lambda = B,$$

(4)

We can get the perspective transformation function expressed as Equation (5) from Equation (4).

$$\lambda = (A^T A)^{-1} A^T B,$$

(5)

If there are n corresponding points, we will get $2n$ equations about perspective transformation function, so we can get the corresponding perspective transformation function through 4 groups corresponding points. Because the perspective transformation is the relationship from pixel coordinates on one plane to another, so we can get different perspective transform coefficients from different front views of different sizes. At the same time, we can get different perspective distortion when we film the same object in a different angle, and the image of front view in different position is different. In order to make the subsequent match precisely intuitive, we should choose the front view of same size, so that we can guarantee the consistent results offering convenience for subsequent image processing.

Using the double array camera capturing the image of the bullet flight, we can obtain the bullet's both sides images when the bullet flies through the adaptive multiple target surfaces. As shown in Figure 3, the left and right view image represents the bullet 1 and bullet 2 fly through the plane and respectively at the same time. When the left camera and the right camera to capture the bullet image, the angle of the camera, which led to the left, and right view image has the projective transformation, so the

collected images are not front view image. The bullets 1 and 2 at the same time t are not in the same plane of the actual three-dimensional space.
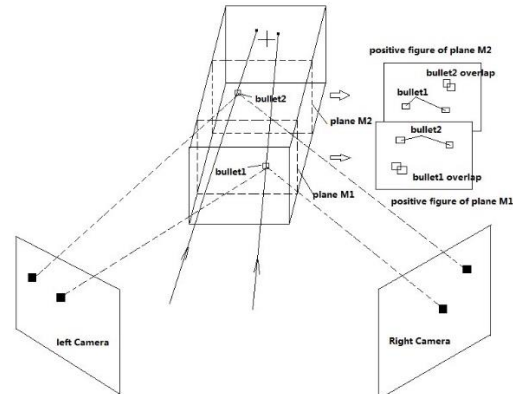
FIGURE 3 The different imaging bullets in the left and right camera

As the same target that the camera of both sides film at the same time must meet the actual space target in one plane of real space. As shown in Figure 3, target 1 is in the adaptive multiple target surface plane M_1 at the moment t. Suppose that the four fixed points are $ABCD$, you'll get these four image point coordinates after image of left and right view. The image point on the left side of these four fixed points in the M plane is marked as A_1, while the right side is marked as A_2. The image point of front view with the uniform size is marked as B. We can obtain the perspective transformation relationship between left view and right view of the uniform size through Equation (5). It can get Equations (6) and (7).

$$A_1\lambda_1 = B,$$

(6)

$$A_2\lambda_2 = B.$$

(7)

Apply λ_1 and λ_2 separately to the left and right view, we can recover the coordinates of bullet point target1_left on the left to the coordinates in the front view and we will get the new image coordinates B_1. When recover the coordinates of bullet point target1_right on the right to the coordinates in the front view, we will get the new image coordinates B_2. Because target1_left and A_1 are in the left view images of M_1 plane, target1_right and A_2 are in the left view image of M_1 plane, and target1_left and target1_right are the images of target1, then we can obtain Equation (8) from Equations (6) and (7).

$$A_1\lambda_1 = A_2\lambda_2.$$

(8)

Then, it can transfer Equation (8) to Equations (9) and (10).

$$A_{t1}\lambda_1 = B_1,$$

(9)

$$A_{t2}\lambda_2 = B_2.$$

(10)

Therefore, we can get that the coordinates of B_1 and B_2 are the same. So we can finish the match of bullet points through finding the intersection plane in which the

image points of front view are fit by multiple iterations. The image is shown in Figure 4.
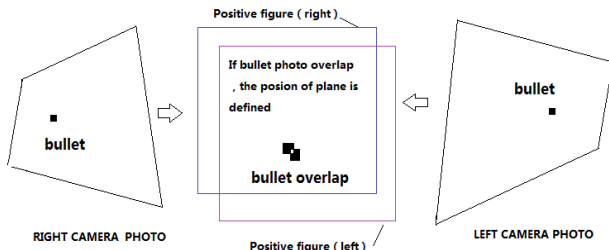
FIGURE 4 Perspective transformation relations of the same bullet

Assume that the two bullet targets are both in M_1 plane. We can respectively recover the two bullet targets to the front view through perspective transformation. The recovery of bullet 1 can be obtained by Equation (9) and Equation (10), at the moment the bullet 1 is in M_1 plane, and we can get the same bullet coordinate points. For the bullet 2, through the same principle analysis, the coordinate of the left image is A_{l11}, while the right is A_{r22}, and the images of front view are C_1 and C_2. They are shown in Equations (11) and (12):

$$A_{l11}\lambda_1 = C_1, \tag{11}$$

$$A_{r22}\lambda_1 = C_2. \tag{12}$$

While in practice, the bullet 2 is not in M_1 plane of the actual space, but in M_2 plane. Thus, the coordinates of C_1 and C_2, which resulted from the transformation of perspective transformation function in the front view can not recover the both sides images to the front view images in M_2 plane where the bullet 2 is. So the coordinates of C_1 and C_2 will not overlap. The recovered front view images of bullet 1 and bullet 2 are shown in Figure 5.
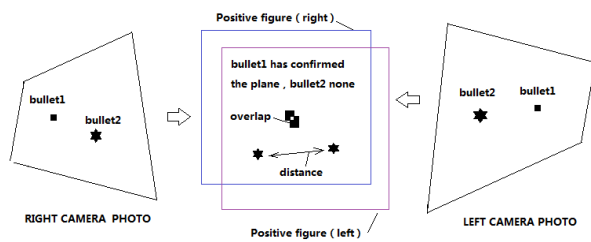
FIGURE 5 Perspective transformation relations of the two bullets

Through the principle of the analysis, we can clear the match of bullet points and achieve the goal of one to one correspondence. According to Equation (13), we can calculate the type *3d* coordinates (X_w, Y_w, Z_w) of space target. The parameters with the letter *l* are calibration parameters and pixel coordinates of the left camera, while the parameters with the letter *r* are calibration parameters and pixel coordinates of the right camera in Equation (13).

$$\begin{bmatrix} u_l m_{l31} - m_{l11} & u_l m_{l32} - m_{l12} & u_l m_{l33} - m_{l13} \\ v_l m_{l31} - m_{l21} & v_l m_{l32} - m_{l22} & v_l m_{l33} - m_{l23} \\ u_r m_{r31} - m_{r11} & u_r m_{r32} - m_{r12} & u_r m_{r33} - m_{r13} \\ v_r m_{r31} - m_{r21} & v_r m_{r32} - m_{r22} & v_r m_{r33} - m_{r23} \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} =$$

$$\begin{bmatrix} m_{l14} - u_l m_{l34} \\ m_{l24} - v_l m_{l34} \\ m_{r14} - u_r m_{r34} \\ m_{r24} - v_r m_{r34} \end{bmatrix} \tag{13}$$

3 Experimental analysis and simulation

3.1 EXPERIMENTAL SCHEME

In order to prove that the proposed algorithm can be implemented with, as well as to solve problems in practical application, experimental simulation standard 5.56mm bullets, high-speed analogue camera with PhantomV12.1. Camera that full frame resolution of 1280×800, full frame shooting rate of 6,242 frames/sec, the maximum recording rate of up to 1,000,000 frames/sec, experimental simulation 6,242 frames/sec.

Given experimental conditions, cannot get PhantomV12.1 speed cameras. So using old-fashioned shoot animation processing forms, taking photos with a digital camera to simulate every artificial speed camera frames of all, according to the bullet velocity 860m/s, each piece can be calculated bullet moving distance of about 0.137 meters, after several manual shooting, and ultimately can get high-speed camera to capture the effect on the bullet. If it does not consider the iris of the camera, capture efficiency and the target, respectively, this and the actual site situation is almost the same, so it can be used to verify the algorithm.

Specific structure of experiments shown in Figure 4, which constitute a rectangle *EFGH* is the target region of the bullet, the trajectory of the bullet 1 and 2. With two thin leads from the target surface area, put a line marker bullet black fake bullet marks, as the bullet frame shooting targets, which two thin three-dimensional trajectory as a follow-up compared in order to test the correctness of trajectory calculation.
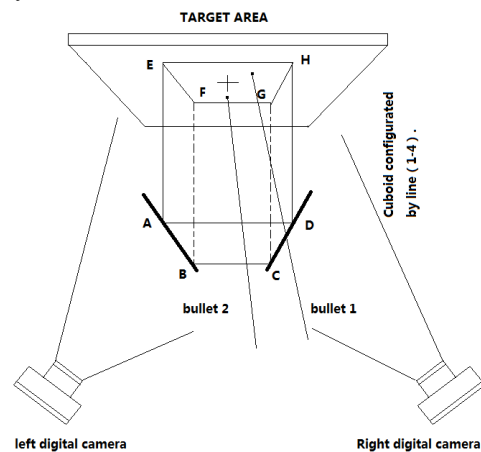
FIGURE 6 Test chart of the algorithm

22

AE, BF, CG, DH are perpendicular to the target surface *EFGH* four thin, with the target area together constitute a virtual rectangular multi- target surface, *A-H* these eight signs point constructed bullet during flight through space plane. Where in the surface and the surface is a surface parallel to each other, the distance between the same. Figure 6 is a schematic structural view of a test algorithm, Figure 7 is a left side of the camera seen from the test structure, and the structure of the right side of the camera as seen in Figure 7 is symmetrical.
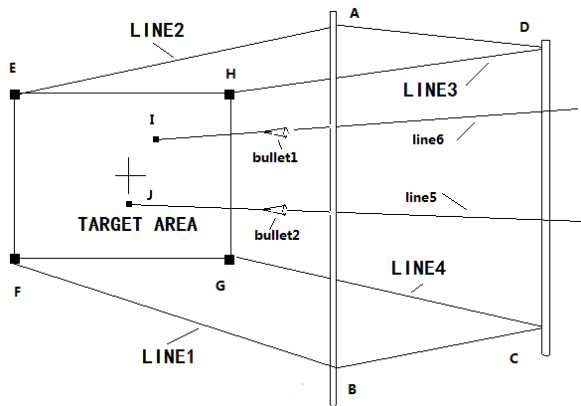
FIGURE 7 Bullet flying through the virtual target surface

This experiment uses a digital camera for two virtual analogue bullets through the process of shooting the target surface, and gets a series of shots left and right view images. Figure 8 and 9 show the left and right camera capture the two bullets images at a time.

FIGURE 8 Left camera to capture the two bullets

FIGURE 9 Right camera to capture the two bullets

3.2 EXPERIMENTAL SIMULATION AND ANALYSIS

Read the collected images in matlab, because the eight landmarks and the two bullets target are small, their coordinates can be manually obtained by selecting the centre of their coordinates. After getting *A-H* the eight landmark coordinates, connect their coordinates in turn. The result is shown in Figures 10 and 11.
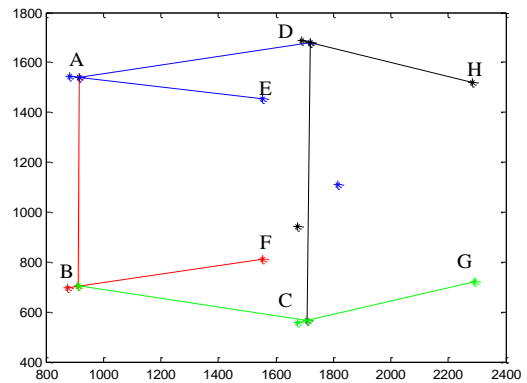
FIGURE 10 Right camera virtual surface composed of eight landmarks
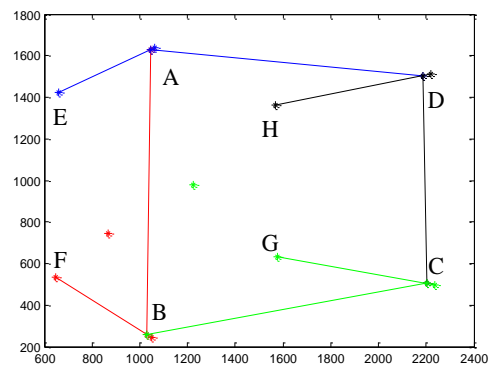
FIGURE 11 Left camera virtual surface composed of eight landmarks

Thus, it can be divided into 20 for *AD, BC, EH; FG* and these points with the same distance are connected in turn, which forming 20 parallel planes. So there is a one-to-one correspondence between the 20 parallel planes and the equidistance 20 planes in three-dimensional space. The result of image segmentation of equidistance is shown in Figures 12 and 13.
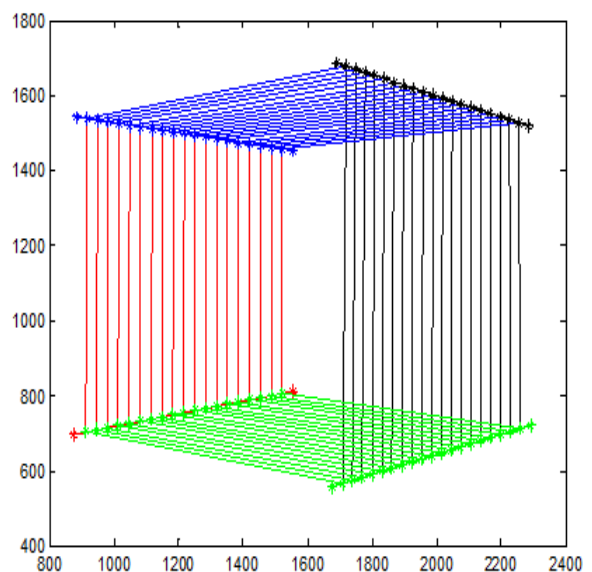
FIGURE 12 The uniform cutting plane of the right view

23

Lei Ming, Chen Mingming, Li Hanshan
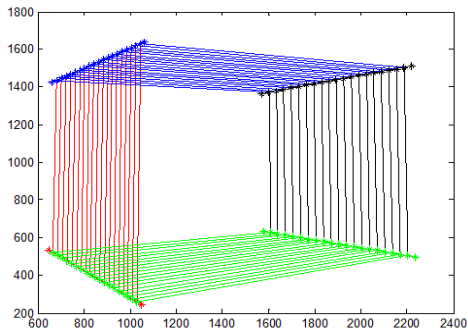
FIGURE 13 The uniform cutting plane of the left view

First assumption, two bullets bullet1 and bullet2 imaging are in the same plane *ABCD* at the same time. By formula it can respectively calculate the perspective transformation relations function about the left, right image plane and the uniform size of the front view. Then, applying the function to the left image and right as the recovery of the bullet point coordinates. Figure 14 shows the two bullets bullet1 and bullet2 in the left view image by perspective transformation to get the two bullets coordinates b1 and b2 in front view image. Figure15 shows the two bullets bullet1 and bullet2 in the right view image by perspective transformation to get the two bullets coordinates b2 and b4 in front view image. By comparing the coordinate size, b1, b2 and b3, b4 the coordinates of the points are not the same. Through the analysis of Equations (11) and (12), the two bullets are not on the plane *ABCD* at the moment, as it not find the bullet into the actual plane at the moment.
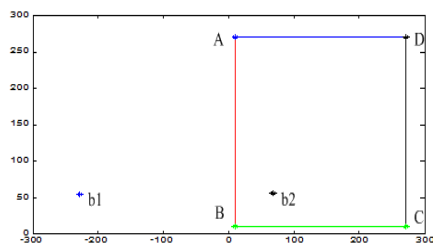
FIGURE 14 The bullets' coordinates b1 and b2 by the left image ABCD conversion
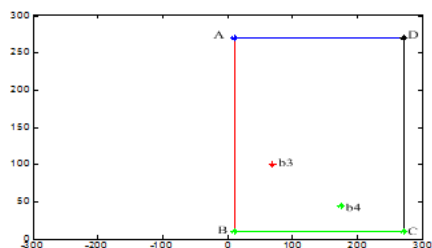
FIGURE 15 The bullets' coordinates b3 and b4 by the right image *ABCD* conversion

Through to the next image plane of the same algorithm, iterative until it finds the coordinate of a bullet is same on the front view image by restoring the left and right view image. Through iteration, it gets a front view of the uniform image plane *MNOP*, and the result is shown in Figures 16 and 17. Figure 16 shows the left view image bullet11 coordinates through transformation to get the coordinates c1 in front view, at the same time, bullet2 coordinate is converted to coordinate c2. Figure 17 shows

the right view image bullet11 coordinates through transformation to get the coordinates c3 in front view, at the same time; bullet2 coordinate is converted to coordinate c4. From the coordinates of the results shows that c1 and c3 is the same, coordinates c2 and c4 are not equal. Through the analysis of Equations (9)-(12), the bullet11 into the plane *MNOP*, the bullet12 is not on the plane *MNOP* at this time, which can be found through the recovery of left and right view the coplanar planes, is to determine the bullet points around the image matching.
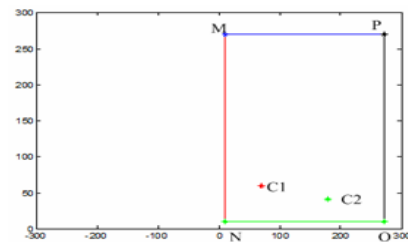
FIGURE 16 The bullets' coordinates C1 and C2 by the left image *MNOP* conversion
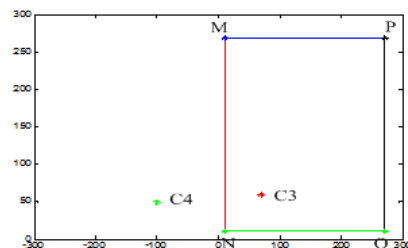
FIGURE 17 The bullets' coordinates c3 and c4 by the right image *MNOP* conversion

Through the above algorithm, multiple iterations around until it finds the coordinates of the bullet is consistent, that is sure to know the same bullet's imaging point in the left and right view. About as long as it can find a set of matching bullets coordinates, then the continuity of the trajectory of the bullet image may be used for matching judgment, then can realize the bullet match in each moment.

Determine the internal and external parameters of camera by camera calibration and combining the Equation (13) that can calculate the bullet's three-dimensional coordinate. Connect the three-dimensional coordinates of the bullet that can get the bullet's three-dimensional trajectory. Figure 18 shows the simulation results of the three-dimensional trajectory.
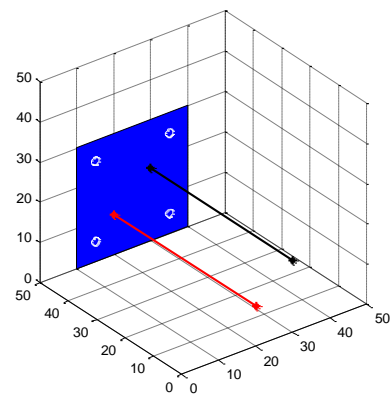
FIGURE 18 Three-dimensional trajectory of the two bullets

4 Conclusions

In order to obtain the flight trajectory of the bullet diagram is the three-dimensional coordinates of different points in time, the need to solve a major problem is that the presence of multi-target, so the camera one by one to the bullet point target pair. This process as the bullet flying through the numerous virtual planes parallel to each other in the process, and the left view image converted by the perspective transformation is a front view of the image. Through multiple iterative methods to find the corresponding coordinates in a front view of time consistent with the target point is about the same time the same target in the viewpoint of a real space in the same plane. This method is an iterative search to find the target by multiple coordinators' consistent elevation view, while the left and right camera will determine the corresponding match point in the bullet. This method is convenient and feasible experimental procedure. The image processing from the surrounding environment can quickly and accurately find the corresponding matching point and thus, can create a three-dimensional bullet trajectories.

References

[1] Rosenfeld A, Johnston E 1973 Angle Detection on digital curves *IEEE Trans Computer* **22** 875-8
[2] Du Z, Lou X 2007 A rapid method of perspective projection correction and spatial location *Journal of Yangtze University* 2007.9 *(in Chinese)*
[3] Hartley R I, Zisserman A 2000 Multiple View Geometry in Computer Vision *Cambridge University Press*
[4] Ling G et al. 2001 Front Matter Multimedia Image and Video Processing *Boca Raton* CRC Press LLC
[5] Werner T, Hersch R D, Hlavac V 1995 Rendering real-world objects using view interpolation *Proceedings of the Fifth International Conference on Computer Vision* 957-62
[6] Maas H G, Green A 1995 Digital photogram metric techniques for high-resolution three-dimensional flow velocity measurements *Optical Engineering* **34**(7) 1970-6
[7] Duda R O, Hart P E 1972 Use of the Hough transformation to detect lines and curves in pictures *Communications of the ACM* **15**(1) 11-5
[8] Casasent D, Krishnapuram R 1987 Detection of target trajectories using the Hough transform *Applied Optics* **26** 247-51
[9] Chen L H 1996 *IEEE Transactions on Image Processing* **5**(3) 547-50
[10] Zhong L, Peide W 1991 Tracking of Maneuvering Targets *National Defense Industry Press* 201-3
[11] Guan S, Fan G, Wang L 2008 Experimental study of the bullet motion simulation *Arrows and guidance* **10**(5) 28

Authors

Ming Lei, born in December, 1970, Xi'an, China

Current position, grades: associate professor, School of Electronic Information Engineering, Xi'an Technological University.
University studies: M.S. degree (2003) on communication engineering at Xidian University.
Scientific interest: image-processing technology, high-speed target image extraction, information-processing technology.
Publications: 16 papers.

Mingming Chen, born in October, 1988, Anhui, China

Current position, grades: Master degree candidate, School of Electronic Information Engineering, Xi'an Technological University.
University studies: M.S. degree (2014) in Electronic Information Engineering at Xi'an Technological University.
Scientific interest: Image processing technology.
Publications: 1 papers.

Hanshan Li, born in July, 1978

Current position, grades: associate professor at the School of Electronic Information Engineering, Xi'an Technological University.
University studies: Doctor's degree (2010) at Northwestern Polytechnical University.
Scientific interest: research and development of photo electricity detection, measurement and control technology, dynamic object test technology, target damage efficiency assessment processing technology.
Publications: 56 papers.

Verification of calculation efficiency of a new CS-PSO algorithm and its application

Juan Yang*, Xuesong Han

Cheng De Petroleum College, Chengde, Hebei, 067000, China

Abstract

Intelligent algorithm is developing rapidly with the development of computer technology. And it is widely used in scientific research and industrial application. As a kind of intelligent algorithm, particle swarm optimization (PSO) has been used in solving problem for a long time. It is based on the bird group behaviour and uses biological group model to find the optimal solution. Its advantages are fast calculation speed and easy implementation while the disadvantages are easily getting into the local extreme, slow convergence speed in the late evolutionary and poor precision. In order to avoid the disadvantages, some modification has been studied for PSO algorithm and establishes the concentration degree and steady degree based PSO (CS-PSO) algorithm in the paper. Based on the convergence performance of particle swarm depends on the particle exploration ability, search space has been adaptively adjusted to improve the convergence performance of particle swarm optimization with the variation of optimal fitness value. Corresponding adjusted method has been shown in the paper. According to the example verification, the CS-PSO is effective and then the algorithm is used in the bellow structure optimization.

Keywords: intelligent algorithm, particle swarm optimization, space adjusted, experimental test

1 Introduction

Intelligent computing is also known as "soft computing", which is affected by natural rules of enlightenment. According to its principle, imitate the algorithm to solve the problems. Principle of bionics design (including the design algorithm) is intelligent computing thought. Intelligent algorithms [1-5] are more and more widely used in solving different optimization problems. Programming of the intelligent algorithm includes linear programming [6-9], dynamic programming [10-12], etc. The content includes many algorithm, such as artificial neural network [13-15], genetic algorithm [16], simulated annealing algorithm [17], and swarm intelligence technology [18].

As an important kind of intelligent algorithm, the PSO is easy to realize with not too many parameters need to be adjusted. It needs not differentiable derivative and related information. PSO can be used as a permutation and combination optimization method to solve mixed integer nonlinear problems [19].

Particle swarm optimization (PSO) algorithm is a new evolutionary algorithm which is developed by Kennedy and Eberhart in 1995. The basic idea is inspired by results of bird group behaviour, and use and improved the biological group model developed by biologists to provide the particles fly to the optimal results in solution space. It belongs to a class of stochastic global optimization technique and finds optimal regions in complex search spaces through the interaction between particles. Its advantages are fast calculation speed and easy implementation while the disadvantages are easily getting

into the local extreme, slow convergence speed in the late evolutionary and poor precision.

In PSO, each solution of the optimization problem is a bird in space, which we call "particle". All the particles have a fitness value, which is determined the optimized function. Each particle has a speed to determine the direction and distance. Then particles follow the current optimal particle to search in the solution space.

PSO is initialized to a group of random particles (stochastic), and then through the iteration to find the optimal solution. In each iteration, particle is updated by following two extreme values. The first is the optimal solution found by particles themselves, and this solution is called. The other extreme is the optimal solution found by all the particles, and this solution is called. It is also just to one part of the particles as neighbour instead of all the particles. In this condition, the optimal solution is the local optimal solution. When particles find the two optimal solutions, they will update the information of themselves.

Iteration termination condition usually select maximum iterative times or current optimal position fit for the minimum adaptive threshold as the terminal conditions. The basic particle swarm optimization will not need the user to determine many parameters, so the method is quite convenient. But it is easy to fall into local optimum, and the search accuracy is not high. Therefore, it is necessary to improve [20-25].

Although there is many improved method for PSO, it still has room to grow. In order to avoid the disadvantages, some modification has been studied for PSO algorithm and establishes the concentration degree and steady degree

---

* *Corresponding author* e-mail: yangjuansky@126.com

based PSO (CS-PSO) algorithm in the paper. Based on the convergence performance of particle swarm depends on the particle exploration capability, search space has been adaptively adjusted to improve the convergence performance of particle swarm optimization with the variation of optimal fitness value. The main contribution of the paper is the proposition of an improved PSO algorithm (CS-PSO). The remainder of the paper is organized as follows: Standard PSO algorithm is described in section 2. CS-PSO method, including some definition and the mathematical model is shown in section 3. Examples verification and engineering application are shown in section 4. Application introduction is shown in section 5. And the conclusion is described in section 6.

2 PSO

2.1 STANDARD PSO algorithm

For arbitrary particle d in the space, position and speed updated at k-th times iteration can be expressed as the x_d^k and v_d^k. The position and speed of $(k+1)$ times of iteration would be updated by the Equations (1) and (2).

$$x_d^{k+1} = x_d^k + v_d^{k+1},\qquad(1)$$

$$v_d^{k+1} = \omega v_d^k + c_1 r_1 (p_d^k - x_d^k) + c_2 r_2 (p_g^k - x_d^k),\qquad(2)$$

where, ω is the inertia coefficient; r_1 and r_2 are the random number between 0 and 1. Generally, $c_1 = c_2 = 2$. Position

vector p_d^k is the optimal position in k-th times iteration of particle d, which is called p_{best}. p_g^k is the optimal position in k-th times of iteration found by all the particles, which is called G_{best}.

The inertia coefficient would vary in the search process with the principle of Equation (3).

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{k_{\max}} \times k .\qquad(3)$$

3 CS-PSO method

How to improve global convergence performance of PSO is always one of the focuses of current research. The convergence performance of particle swarm depends on the particle exploration capability in the solution space. The global convergence performance of PSO algorithm can be improved by improving the particles search ability.

3.1 DEFINITION

In this paper, we will give some definitions and these definitions would be used in the modified PSO method.

Definition 1: Set $p_{best}(i,k)$ as the fitness value for the best position of the particle i at time k, and set $g_{best}(k)$ as the best position of all particle's fitness value. Then we define the concentration degree $con\_\deg(k)$ as the following:

$$con\_\deg(k) = \begin{cases} \dfrac{|g_{best}(k)|}{\dfrac{1}{M}\sum\limits_{i-1}^{M}|p_{best}(i,k)|} & \dfrac{1}{M}\sum\limits_{i-1}^{M}|p_{best}(i,k)| \neq 0 \\[4ex] \lim \dfrac{|g_{best}(k)|}{\dfrac{1}{M}\sum\limits_{i-1}^{M}|p_{best}(i,k)|} & \dfrac{1}{M}\sum\limits_{i-1}^{M}|p_{best}(i,k)| = 0 \end{cases}.\qquad(4)$$

Apparently, $con\_\deg(k)$ is between 0 and 1. After iteration for a certain time, $con\_\deg(k)$ would move close to 1. This means the particles concentrate together.

$con\_\deg(k)$ reflects all particle concentration degree at time k.

Definition 2: Set the steady degree $ste\_\deg(k)$ as:

$$ste\_\deg(k) = \begin{cases} \dfrac{|g_{best}(k)|}{\dfrac{1}{S}\sum\limits_{i-1}^{S}|g_{best}(i,k)|} & \dfrac{1}{M}\sum\limits_{i-1}^{M}|p_{best}(i,k)| \neq 0 \\[4ex] \lim \dfrac{|g_{best}(k)|}{\dfrac{1}{S}\sum\limits_{i-1}^{S}|p_{best}(i,k)|} & \dfrac{1}{S}\sum\limits_{i-1}^{S}|p_{best}(i,k)| = 0 \end{cases}.\qquad(5)$$

In the definition, S is the smooth order. It usually uses an integer between 10 and 200. $ste\_\deg(k)$ is also between 0 and 1. After iteration for a certain time, the $ste\_\deg(k)$ would move close to 1. Then, the velocity of the particle is approximately to 0. It reflects the steady degree of the particle at the time k.

Obviously, at a certain time in the iteration, all or part of the particles can be modified to jump out of local solutions to obtain the global optimal solutions with the concentration and stable degree of the particle.

3.2 CONCENTRATION AND STEADY DEGREE BASED PSO ALGORITHM (CS-PSO)

Searching in the standard PSO algorithm, particle swarm will move to the local minimum or global minimum convergence. In this case, PSO often has searched the region or close region with global optimal position in it. Then, the searching speed is much slower than earlier period, the search speed is almost zero and the local search ability is extremely low. Therefore, it is necessary to avoid

the occurrence of this phenomenon by effective measures to provide finer search for the small range of the global optimal solution to improve the efficiency and accuracy. In this condition, search space should be adaptively adjusted to improve the convergence performance of particle swarm optimization with the variation of optimal fitness value.

Main idea of the CS-PSO.

Set $g_{best}(k)$ as the fitness value of particle swarm optimization at the iteration step K, the corresponding optimal position is $x_g = (x_{g1}, x_{g2}, ..., x_{gN})$. The original variation interval of dth dimension variable particle is $region(d) = [s_{left-d}, s_{right-d}]$, $d = 1, 2, ..., N$. Shrinkage factor of the interval $\alpha \in (0,1)$ and maximum permissible error of the interval is δ. After iteration for K steps, then $g_{best}(k) - g_{best}(k-L) < \varepsilon$, and $K - L > 0$, ε is the permissible error with given fitness. L is a positive integer. Adjustment of particle swarm search space will be given according to the following formula:

$$region(d) = \begin{cases} [x_{gd} - \alpha(1-\alpha)(x_{gd} - s_{left-d}), x_{gd} + \alpha(1-\alpha)(s_{right-d} - x_{gd})] \\ if \max\{s_{rght-d} - s_{left-d}\} > \delta \\ [s_{left-d}, s_{right-d}] \\ others \end{cases} , \quad d = 1, 2, ..., N . \quad (6)$$

So, as the iteration proceeds, convergence speed of particle swarm optimization ability and local exploration would be enhanced according to the contraction of particle swarm search space to improve the convergence accuracy of particle swarm. When it at a certain moment and the solution space contraction to a certain extent, it can be extended particle swarm search space to provide all particles have fully variation in the new space to jump out of local optimum to obtain the global optimum. This method can balance the efficiency and precision of the optimization process"

The CS-PSO algorithm can be divided in to some steps as shown in Figure 1.

4 Verification

4.1 MATHEMATICS MODEL VERIFICATION

Parameters in the PSO will be set as: the inertia weight of the maximum and minimum values is 0.9 and 0.1 respectively. Maximum iteration number is 2000 and independent operation for 50 times.

1) Test function is shown as the following:

$$\min f_1(x) = (x_1 - 10)^3 + (x_2 - 20)^3,$$

$$\text{S.t} \begin{cases} 100 - (x_1 - 5)^2 - (x_2 - 5)^2 \le 0 \\ (x_1 - 6)^2 + (x_2 - 5)^2 - 82.81 \le 0 \\ 13 \le x_1 \le 100, \quad 0 \le x_2 \le 100 \end{cases},$$

The global optimal solution is $x^* = (13.985, 0.8216)$, and then $f(x^*) = -6908.324$.
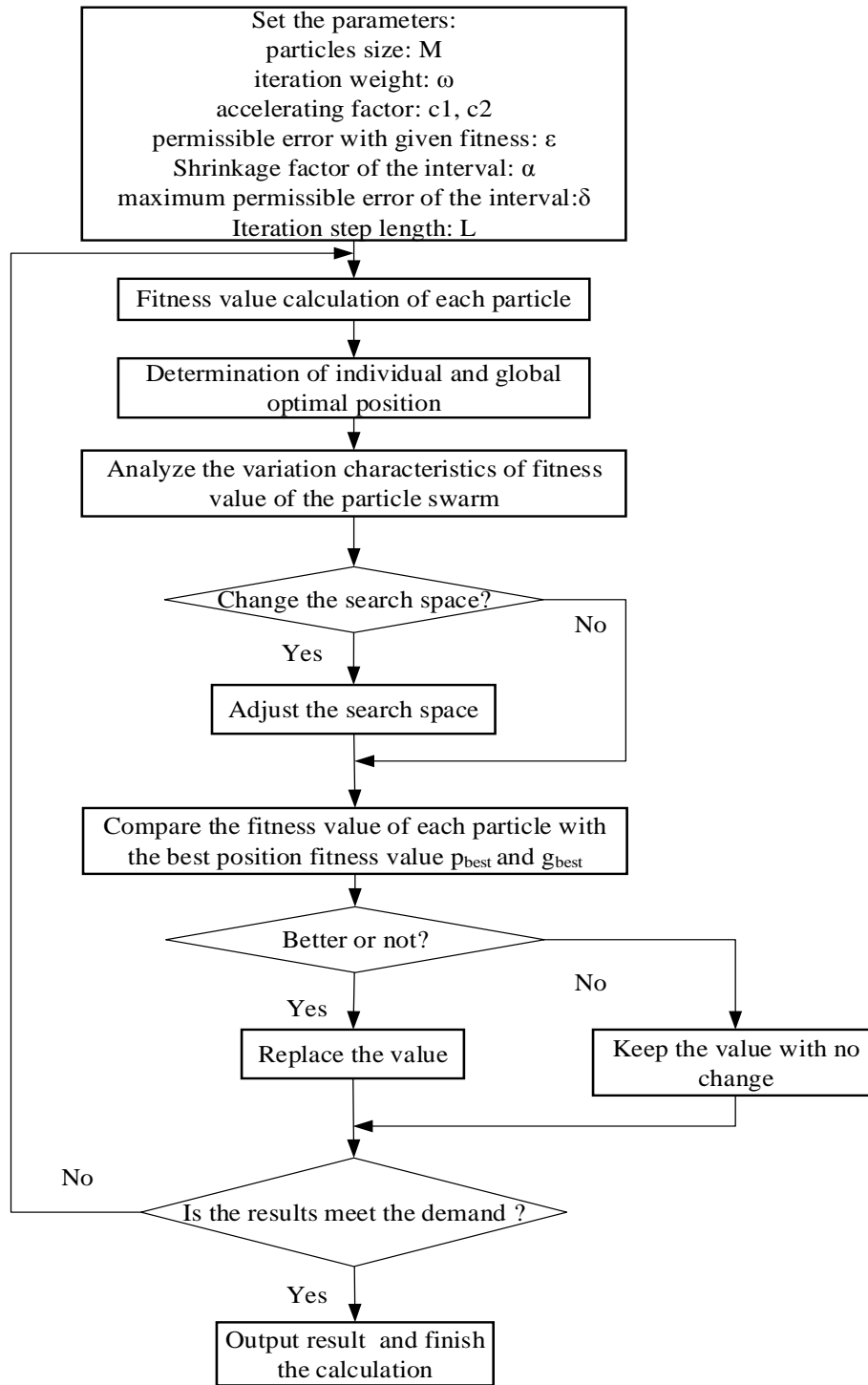
FIGURE 1 Flow chart of CS-PSO

2) Test function is shown as the following:

$$\min f_1(x) = (x_1 - 10)^3 + (x_2 - 20)^3.$$

$$\text{S.t.} \begin{cases} 0 \le 85.125 + 0.005723 x_2 x_5 + 0.000624 x_1 x_4 - 0.002204 x_3 x_5 \le 100 \\ 90 \le 80.5125 + 0.007231 x_2 x_5 + 0.0030032 x_1 x_2 + 0.0021813 x_3^2 \le 110 \\ 25 \le 9.300736 + 0.0047012 x_3 x_5 + 0.0012591 x_1 x_3 - 0.00190021 x_3 x_4 \le 25 \\ 75 \le x_1 \le 103, \ 33 \le x_2 \le 50 \quad 25 \le x_i \le 45, \ i = 3, 4, 5 \end{cases}.$$

The optimal value is $f(x^*) = 680$.

3) $\min f_3(x) = x_1^2 + (x_2 - 1)^2$.

S.t $\begin{cases} x_2 - x_1^2 = 0 \\ -1 \le x_i \le 1, \quad i = 1,2 \end{cases}$.

The optimal value is $f(x^*) = 0.74$.

4) $\min f_4(x) = x_1^2 + 4x_2^2$.

S.t $\begin{cases} x_1 - x_2 - 1 \le 0 \\ x_2 - 1 \le 0 \\ -x_1 - x_2 + 1 \le 0 \end{cases}$.

The optimal value is $f(x^*) = 0.800001$, while the $x^* = (0.8, 0.2)$.

Table 1 gives the results comparison of algorithm proposed in the paper and other algorithms referred in other paper. In the table, results of algorithm proposed in the paper show better performance than PSO.H and GA algorithms. The optimal value found by the particle size of 20 with the method proposed in the paper has a good performance with the optimal value found by particle size of 20 with the other method. For some functions, algorithm proposed here can get better results.

TABLE 1 Results comparison with other algorithms

Function	TRPSO		$f(x^*)$ M=20; $c_1=c_2=1.8$	PSO.H	GA
	Particles size M=40; $c_1=c_2=1$	M=20; $c_1=c_2=1.8$			
$f_1(x)$	-6961.813876	-6961.813876	-6961.813876	-6961.7	-6952.1
$f_2(x)$	-30665.538672	-30665.538672	-30665.538672	-30665.5	-30664.5
$f_2(x)$	0.749912	0.749917	0.749990	0.75	0.75

TABLE 2 Results comparison with other algorithms

Function		Number of experiments	Effect number of experiments	Mean number of iteration
$f_3(x)$	TRPSO	50	50	12.1
	MRR	10	6	30
$f_4(x)$	TRPSO	50	50	16.3
	MRR	10	10	25

The algorithm in calculation is compared with the other algorithms by the convergence rate and stability. The results are shown in Table 2. It is shown that the convergence efficiency and velocity of the algorithm in this paper are higher. The experimental results show the effectiveness and rationality of the proposed algorithm, which means certain advantage of the algorithm.

4.2 ENGINEERING EXAMPLES

Bellows expansion joints are widely used in petroleum system, chemical plant and the nuclear system. They are the temperature and stress displacement compensation element. The bellows are the main components, and the structural design parameters are integer and discrete. The design process involves many performance constraints and highly nonlinear for objective function. Bellow optimization design problem belongs to problems with nonlinear constrained discrete design variables.

1) Use this method to optimize the structure parameters for bellow with design pressure of 0.25MPa and diameter of 400 mm. The optimal solution of $x_1, x_2, x_3, x_4(n, N, h, t)$ is found the least unit volume weight.

x_1, x_2, x_3, x_4 are layers number, wave number, wave height and single wall thickness. x_1, x_2, x_3, x_4 are all discrete variables, and x_1, x_2, x_3 are integer multiple while x_4 is the integer multiple of 0.2.

The constraint condition of the self-variable are:
$\begin{cases} 1 \le x_1 \le 10 \\ 1 \le x_2 \le 10 \\ 30 \le x \le 70 \\ 0.6 \le x_4 \le 1.2 \end{cases}$.

Corrugated pipe material is 316L and the design temperature is 300℃. The pitch of waves is $q = 41mm$, and fatigue life cycles is 10500 times. Taken steel plate utilization into consideration, the bellows expansion length should be 500 mm with an error less than 5%.

Constraint conditions of calculation method, relevant parameters, strength, stability and fatigue life of the bellows are referred with related papers. For the simplification of calculation, strength factor of $C_m = 3.0$ are adopted. Then, the mathematics model for the bellow can be described as the following:

Find $X(x_1, x_2, x_3, x_4)$ to provide the $\min f(x) = \dfrac{Q}{E}$,

$Q = (D_b + x_1 x_3) \times 3.14 x_1 x_3 \rho$, $L = (2x_3 + 0.57lq)x_2 + 2L_l$, where, ρ is the density of bellow material; D_b is the diameter.

$Q = (D_b + x_1 x_3) \times 3.14 x_1 x_3 \rho$, $E = x_2(e - 0.15q\phi)$, where, e is the compensation rate for bellow.

$$e = \max\left\{ e_1 = e_\theta + e_y - |e_x|, e_2 = e_\theta + \frac{0.15q\phi}{c\theta} - |e_x| \right\},$$

$$e_{x,y} = \frac{[2\sigma_T - 0.7(\sigma_T + \sigma_T)]x_3^2}{\dfrac{E_b t_p^2}{2x_3 C_f} + \dfrac{5E_b t_p}{3C_d}}, \text{ where, } t_p \text{ thickness of the}$$

bellow; e_x, e_y, e_θ are the displacement of axial, horizontal and angular direction; C_d, C_f, C_θ, ϕ are the coefficients. E_b is the elastic modulus which is subject to $\begin{cases} p - p_{sc} \le 0 \\ p - p_{si} \le 0 \end{cases}$, where, p is the design pressure; p_{sc}, p_{si} are column instability and plane instability pressure limit.

$$\begin{cases} \sigma_1 - \sigma'_{sb} \le 0 \\ \sigma_2 - \sigma'_{sb} \le 0 \\ (\sigma_3 + \sigma_4) - 3\sigma'_{ab} \le 0 \end{cases}, \text{ where, } \sigma'_{sb} \text{ is the allowable}$$

stress for the corrugated pipe with design temperatures; $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ are the film pressure and bending stress due to the stress for corrugated pipe, respectively.

$0.95L_d \le L \le L_d$, $N_C - [N_C] \le 0$, where, L_d is expansion to fixed length or their divisibility times, L is the bellow expansion length, $[N_C]$ is the allowed design life of bellow, N_C is the working life of bellow.

The equations described above represent the variable constraint, stability constraint, strong constraint, expansion length constraint and fatigue life constraints.

2) The optimization calculation and result analysis

To meet the requirement of engineering, the optimal solution should be discrete solution in the feasible domain. Define initial constraint penalty factor is big enough relative to the optimal objective function. With the test results, define $r = 100$, discrete penalty factor $s_{d0} = 1$ and $c = 1.005$. The maximum iterations number $K_{\max} = 100$. Particle number $N_d = 50$ and $\omega_{\max} = 1.2$, $\omega_{\min} = 0.4$. Optimization design of discrete variables is processed for the bellow and calculation results and theoretical solution is included in the Table 3 by using the grid method.

TABLE 3 Comparison of results with different calculation methods

Method	x_1	x_2	x_3	x_4	$F(x)$	$s\phi(x)$	$rG(x)$	$\dfrac{Q}{E}$
Existing parameters	1	3.0000	45	1				0.086
New method	1	4.9893	30	0.58	0.046	0	0	0.0472
Theoretical solution	1	5.0000	32	0.6				0.047

From the result of this method, the optimization target value increases by 66.31%. The difference between the method and the theoretical solution is less than 1%. The optimal solution is $(1, 4.9893, 30, 0.6)$. The discrete degree is very high and satisfies the discrete accuracy requirements.

In the solving process, 10 local optimal solutions have been acquired. It can be seen that when it is at iteration times of 29, the optimal solution is approximately to the global optimal solution (Figure 2).
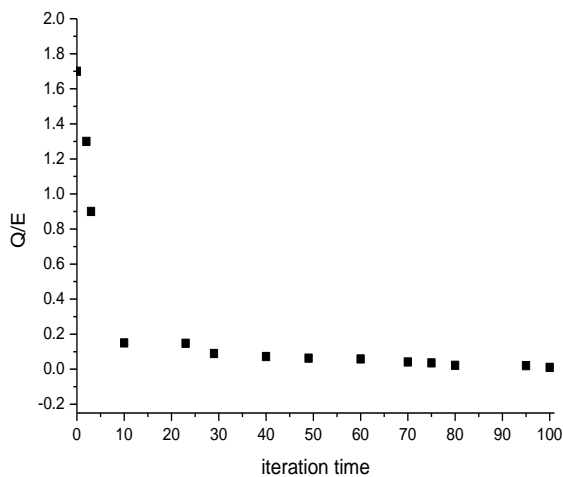
FIGURE 2 Local optimal solution

5 Application introductions

PSO algorithm is widely used in neural network training for function optimization, fuzzy system control and other intelligent algorithm. Here we give some introduction of the applications of the algorithm.

5.1 OPTIMIZATION OF FUNCTION

Study of particle swarm algorithm and convergence is to study and solve the optimization problem much better. Usually, this kind of problem is quite complex for the problem having character of large scale and high dimension. At the same time, in the mathematical nature, the complexity can be reflected as non-linear, non-convex and non-differentiable calculus properties and large number of local extreme values exists in function distribution. Compared to the traditional deterministic optimization algorithm, the PSO algorithm has the characteristics of fast reaction and high sensitivity in solving this kind of problem with the proper choice of initial value. Other global optimization algorithms, such as genetic algorithm, simulated annealing algorithm, evolutionary programming, etc. have different mechanisms and single structure, and it is difficult to achieve efficient optimization in complex functions with

high dimension and direction. But the PSO algorithm combines the advantages and disadvantages and it can achieve efficient optimization of this kind of problems.

5.2 NEURAL NETWORK TRAINING

The PSO algorithm is now mainly exist in 3 aspects of training neural network:
1) network topological structure and function transfer;
2) weight settings in connection;
3) intelligent learning. Each particle can completely describe all the related parameters of the neural network. Optimization of these parameters can be obtained with repeated updating to achieve the training effect. Compared with similar types of learning algorithms, such as back propagation algorithm, particle swarm algorithm used in the neural network has advantages of without the help of differentiable derivative and differential properties in transferring information between the functions. PSO also can get better results than the error back-propagation algorithm with higher speed under high probability conditions.

5.3 PARAMETERS OPTIMIZATION

Particle swarm optimization algorithm can be used in parameters optimization in continuous and discrete problems. The problems mainly include signal processing, path planning of robot, the fuzzy controller design, and the pattern recognition problem

5.4 COMBINATORIAL OPTIMIZATION

Particle swarm algorithm with "01" string coding needs a more reasonable solution and measures in many ordered structure problems in combinatorial optimization problem and the expression of constraint handling problems. When the problems are different, correspond particles

descriptions are different. It can be solved by redefining a new operator. The optimization scheme has solved a variety of TSP, VRP and scheduling problem in factories.

6 Conclusions

Intelligent algorithms are more and more widely used in solving different optimization problems. As an important kind of intelligent algorithm, the PSO is easy to use for the reason of not too many parameters to be changed. Particle swarm optimization (PSO) algorithm belongs to a class of stochastic global optimization technique and finds optimal regions in complex search spaces through the interaction between particles.

In order to avoid the disadvantages, some modification has been studied for PSO algorithm and establishes the concentration degree and steady degree based PSO (CS-PSO) algorithm in the paper. From the convergence performance of particle swarm depends on the particle exploration capability, search space has been adaptively adjusted to improve the convergence performance of particle swarm optimization with the variation of optimal fitness value. Corresponding adjusted method has been shown in the paper.

The method is applied in some functions to verify the validity, and then bellows structure design has been processed with the CS-PSO algorithm. The design process involves many performance constraints and highly nonlinear for objective function. Bellow optimization design problem belongs to problems with nonlinear constrained discrete design variables. All the application shows the availability of the method.

Although improved algorithm improves the global search ability and convergence performance of PSO algorithm to a certain extent and the significant effect of various control parameters on the performance of the proposed algorithms. How to balance control parameters in the algorithm is still needed to further studied.

References

[1] Zhang Q, Li X, Tran Q 2003 Particle swarm optimization based hybrid intelligent algorithm *International Conference on Machine Learning and Cybernetics* **3** 1648-50

[2] Wan S, Sun L 2009 Hybrid intelligent algorithm for optimization design of annular water supply network *2009 International Workshop on Intelligent Systems and Applications* ISA 2009 1-4

[3] Zhao R, Song K 2003 A hybrid intelligent algorithm for reliability optimization problems IEEE International Conference on Fuzzy Systems **2** 1476-81

[4] Zhao M 1997 Hybrid computational-intelligent algorithm for the nonlinear least-squares global solution *Journal of Software* **8**(7) 555-60

[5] Zhang Y, Chen S, Wan Y 2009 An intelligent algorithm based on grid searching and cross validation and its application in population analysis *Proceedings of the 2009 International Conference on Computational Intelligence and Natural Computing CINC 2009* **2** 96-9

[6] Punekar M, Flanagan M F 2010 Low complexity linear programming decoding of nonbinary linear codes *The 48th Annual Allerton Conference on Communication, Control, and Computing* Allerton 2010 6-13

[7] Zhang X, Siegel P H 2012 Adaptive cut generation algorithm for improved linear programming decoding of binary linear codes *IEEE Transactions on Information Theory* **58**(10) 6581-94

[8] Wan S-P, Dong J-Y 2014 Possibility linear programming with trapezoidal fuzzy numbers *Applied Mathematical Modelling* **38**(5-6) 1660-72

[9] Gordon G J, Hong S A, Dudík M 2009 First-order mixed integer linear programming *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence UAI 2009* 213-222

[10] Liu F, Sun J, Si J, Guo W, Mei S 2012 A boundedness result for the direct heuristic dynamic programming *Neural Networks* **32** 229-235

[11] Zhou Y, Cao L, Pan Y 2009 The optimal inventory model based on the dynamic programming method. *Proceedings of the 2nd International Conference on Modelling and Simulation ICMS2009* **3** 74-8

[12] Li W, Xu G, Wang Z, Xu Y 2008 Dynamic energy management for hybrid electric vehicle based on approximate dynamic programming *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)* 7859-63

[13] Zhang Z, Gao J, Zhong Q, Li S 1997 Application of artificial neural network on diagnosis of cold brittle fracture *Journal of Materials Engineering* **5** 28-30

[14] Xu W, Wang D, Zhou Z, Chen H 1997 Application of artificial neural network combined by genetic algorithm in fault diagnosis of power transformer *Proceedings of the Chinese Society of Electrical Engineering* **17**(2) 109-12

[15] Sun C, Bao J, Zhang J, Huang J 2003 Study of artificial neural network applied in the active control of diesel vibration *Journal of Shanghai Jiaotong University* **37**(5) 781-4

[16] Vivekanandan P, Rajalakshmi M, Nedunchezhian R 2013 An intelligent genetic algorithm for mining classification rules in large datasets *Computing and Informatics*, **32**(1) 1-22

[17] Han W, Jin X, Ren H 2007 Defect reconstruction from MFL signals using genetic-simulated-annealing algorithm *Journal of Basic Science and Engineering* **15**(2) 257-62

[18] Cao L, Jin Q, Fan T, Su W 2012 On-line identification of process model based on swarm intelligence technology *Advanced Materials Research* **403-408** 3216-9

[19] Eberhart R, Kennedy J A 1995 New Optimizer Using Particle Swarm Theory *Proceedings of the Sixth International Symposium on Micro Machine and Human Science* 39-43

[20] Ding W, Wang J, Guan Z 2012 Cooperative extended rough attribute reduction algorithm based on improved PSO *Journal of Systems Engineering and Electronics* **23**(1) 160-6.

[21] Gu T-Y, Ju S-G, Han Fei 2009 An Improved PSO algorithm encoding a priori information for nonlinear approximation *ICIC'09 Proceedings of the Intelligent computing 5th international conference on Emerging intelligent computing technology and applications* 223-231

[22] Chen Q, Guo W, Li C 2009 An improved PSO algorithm to optimize BP neural network *The 5th International Conference on Natural Computation ICNC 2009* **2** 357-60

[23] Liu Z, Li L 2013 An algorithm of license plate recognition based on improved PSO-BP network *Mechanics and Materials* **427-429** 1714-7

[24] Chen Y, Sun H, Wen J, Cheng S, Huang J 2011 Application of improved PSO algorithm in network reconfiguration of shipboard power system *Electric Power Automation Equipment* **31**(3) 29-34

[25] Thanushkodi K, Deeba K 2012 On performance analysis of hybrid intelligent algorithms (Improved PSO with SA and Improved PSO with AIS) with GA, PSO for multiprocessor job scheduling *WSEAS Transactions on Computers* **11**(5) 131-47

Authors

Juan Yang, born in May, 1979, Shuangyashan, Heilongjiang, China

Current position, grades: lecturer.
University studies: Chengde petroleum college.
Scientific interest: wireless sensor network communication and information fusion.

Xuesong Han, born in July, 1980, Jilin, Jilin, China

Current position, grades: lecturer.
University studies: Chengde petroleum college.
Scientific interest: wireless sensor network communication and information fusion.

Structure optimization of cycloid gear based on the finite element method

Jianmin Xu[1], Lizhi Gu[2*], Shanming Luo[3]

[1]*College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China*

[2]*College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China*

[3]*School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, China*

Abstract

A numerical model of cycloidal gear is created by using three-dimensional software and finite element analysis is applied with ANSYS platform. The first six natural frequencies and mode shapes are obtained, as a result. Influences from structure, material and thickness of the gear are investigated. Analysis shows that, modal shapes of cycloid gear are mainly circumferential modes, umbrella-type modes, torsional vibration mode and radial modes. The first six natural frequencies of 5 kinds of cycloid gear with variable cross-section were smaller than those of ordinary cycloid gears, and cycloid gears with variable cross-section can avoid resonance frequencies easily. Dynamics of five new cycloidal gears with variable cross-sections are consistent with ordinary cycloid gears. Modal frequencies of ordinary cycloid gears increases in accordance with materials, such as bearing steel, alloy steel and plastics; also, natural frequency increases with the increase of the thickness of the gear. Conclusions of this paper provide a basis for dynamic designing of cycloid gears.

Keywords: cycloid gear, finite element modal analysis, free modal, constraint modal

1 Introduction

In recent decades, many scholars have done a lot of researches in terms of cycloid drive. Yan Hong Sen and Ta Shi Lai [1] established a basic planetary gear system based on a cylindrical tooth profile. Laita-Shi [2-4] proposed a mathematical model and design procedures for planetary transmission systems with cycloid gears and derived the equation of meshing cycloid drive from conjugate surface theory. Leilei et al [5] built a finite element model of the cycloid gear and needle teeth based on ANSYS software and conducted three-dimensional contact analysis. Thube S. V and TR Bobak [6] established a finite element model of the rotating parts of the cycloid reducer and conducted a dynamic analysis. Zhang Xiu Yan and Xiao Jun Dai [7-8] designed a new type of tetracyclic cycloid gear, conducted meshing stiffness analysis and finite element modal analysis based on the finite element software. Biernacki and Krzysztof [9] designed a plastic cycloidal gear and conducted its finite element analysis. Nam, WK, JW Shin, and SH Oh [10] designed a thin ball reducer for a robot. Xiaojun Jun and Wei-Dong [11] established solid model of RV cycloid reducer using Pro/E software and built dynamic analysis models, analysed the inherent characteristics of cycloid using ANSYS software. Liji Shun et al [12] established solid model of cycloid gear with two teeth difference using Pro/E software and analysed the inherent characteristics of cycloid including non-binding modes and modal with actual boundary constraints using

ANSYS software. Qin Guang Yue et al [13] conducted modal analysis for Shearer cycloidal gear based on ANSYS software. Woody and Zhang you Chen [14] established cycloid gear model for multiple teeth difference using ANSYS analysis software, derived root stress calculation formula of cycloid gear with different meshing phase angles and calculated and analysed root stress of cycloid gear with different meshing phase angles.

Cycloid planetary drive is a drive with broad application prospects. Cycloid planetary transmission has significant advantages including transmission ratio range, hardened (bearing steel) multiple-gear meshing, small, smooth motion, long life, low noise, high load capacity and high transmission efficiency. Cycloid planetary transmission has been widely used in many industries and occupies a very large portion in reducer industry. Tooth profile of cycloid gear is composed of hypocycloidal or epicycloid and is a key component of cycloidal gear reducer. As a part of the main transmission, cycloid gear bears the coupling of multiple parts in the meshing process. If its own natural frequency coincides with the drive frequency, it may cause forced vibration of whole reducer. So modal frequencies for each cycloid is calculated in this paper, it can provide a theoretical basis to avoid resonance and the emergence of harmful vibration mode and study dynamics in depth.

---

* *Corresponding author* e-mail: gulizhi888@163.com

2 Mathematical model of tooth profile for an ordinary cycloid gear

One method of forming cycloidal tooth profile is of formation with two external circles, another method is of formation with two circles inner meshing. Formation method with two external circles meshing is described here. The two circles are named a roll circle and a base circle, respectively (insert a figure, Figure 1). When the roll circle and the base circle are tangent externally, and the roll circle makes pure rolling against the base circle, the trajectory of any fixed point on the roll circle is a epicycloid. Equations of a cycloid gear in Cartesian coordinate system are as follows.

$$
\begin{aligned}
X_1 &= (r_1 + r_2)\sin\phi - r_2\sin\left(\frac{r_1}{r_2}\phi\right), \\
Y_1 &= (r_1 + r_2)\cos\phi - r_2\cos\left(\frac{r_1}{r_2}\phi\right),
\end{aligned} \tag{1}
$$

where r_1 represents the radius of the base circle; r_2 represents roll circle radius; ϕ represents the angle of the roll circle rotates about the centre of the base circle in forming a cycloid.

3 Finite element modal analysis for ordinary cycloid gear

3.1 MODAL ANALYSIS THEORY

A modal is the representation of natural vibration characteristics for a mechanical structure, each modal has specific natural frequencies and mode shapes. Modal analysis in this paper is used to obtain the natural frequencies of cycloid gears. Free vibration equation is as follows:

$$
[M]\{\ddot{u}\} + [K]\{u\} = \{0\}, \tag{2}
$$

where $[K]$ is stiffness matrix, $[M]$ is mass matrix, $\{\ddot{u}\}$ and $\{u\}$ are acceleration vector and displacement vector respectively. Algebraic equation of natural frequency is obtained by solving free vibration equation.

$$
w^{2n} + a_1 w^{2(n-1)} + \dots + a_n w^{2n} + a_n = 0. \tag{3}
$$

Order natural frequencies and mode shapes of the structure can be obtained by solving Equation (3).

3.2 ESTABLISH FINITE ELEMENT MODEL OF ORDINARY CYCLOID GEAR

Cycloid gear geometric model is established with 3D modelling software, as shown in Figure 1. The 3D model is imported to ANSYS and finite element modal analysis is conducted. Parameters and procedures are as follows.

1) Modulus of elasticity is 2.07×10^5 Mpa, Poisson's ratio is 0.25 and the density is 7.8×10^3 kg/m^3.

2) Model meshing is conducted. Mesh quality and density have a very big impact on the results of finite element analysis and have direct impact on the quality of the final grid results. Because cycloidal tooth profile is complex, the sweep mesh partition method is adopted in this paper. The model grid is mainly Hexahedral elements. Final model is divided into 107,955 units and has a total of 162,553 nodes. Finite element model of ordinary cycloid gear is shown in Figure 2.

3) Analysis type is specified as modal analysis.

4) Set number of extended modal. Structure vibrating can be expressed as linear combination of vibration modes corresponding to natural frequencies. Low order modes have great impact on the vibrational structure, low natural frequencies and mode shapes have practical significance. So generally the first 5 to 10 modal analysis are performed; in this paper, the first six natural frequencies and mode shapes are calculated.
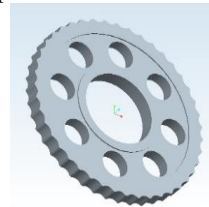
FIGURE 1 Geometry of ordinary cycloid gear
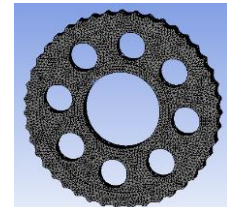
FIGURE 2 Finite element model of the ordinary cycloid gear

3.3 RESULTS ANALYSIS

According to modal analysis theory, modal characteristics under arbitrary boundary constraint conditions can be calculated by mathematical modelling methods from modal parameters calculated under free boundary conditions. On the contrary, the results obtained under the specified boundary conditions cannot be converted to the dynamic characteristics of other boundary constraints. It is necessary to analyse the differences for ordinary cycloid gear between modal analysis under free boundary conditions and modal analysis under actual boundary constraints. Figure 3 shows the comparison of natural frequency for cycloid gear between free modal and constrained modal. As can be seen, the first-order and second-order natural frequency of ordinary cycloid gear under constraint modal is bigger than that of free modal. But the third-order to sixth-order natural frequency of ordinary cycloid gear under constraint modal are smaller than that of free modal.
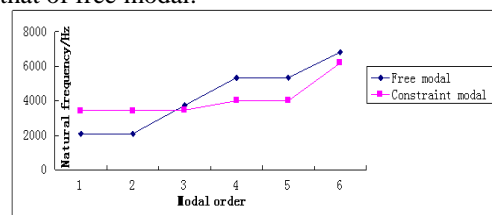
FIGURE 3 The comparison of natura3l frequency for ordinary cycloid gear between free modal and constraint modal

4 Structure impact

4.1 DESIGN FIVE NEW VARIABLE CROSS SECTION CYCLOIDAL GEAR

Cycloidal gear with variable cross-section is obtained when tooth profile changes along the axis of the gear, linearly or non-linearly. Each cross-section of variable cross-section cycloid gear is a tooth profile of ordinary cycloid gear. Five kinds of variable cross-section cycloid gear are designed and 3D structure is shown in Figure 4. These shapes include concave, drum-shaped, spherical, oblique and tapered. The outer periphery tooth surface of concave cycloidal gear (Figure 4a) is the concave surface, corresponding needle teeth are convex central drum. Central convex drum-shaped surface of needle teeth mesh with the outer periphery concave surface of cycloid gear. Drum-shaped cycloid gear (Figure 4b) is contrary to the cycloidal gear with concave structure. Its outer circumferential tooth surface of cycloid gear is convex central drum surface, corresponding needle teeth are concave surface. Needle teeth of spherical cycloid gear (Figure 4c) is spherical, corresponding outer periphery of cycloid gear is arc-shaped surface meshing with the ball needle teeth. Needle teeth of oblique cycloidal gear (Figure 4d) are inclined cylinder, their outer periphery teeth of cycloid are inclined surface. The tooth profile of Conical cycloidal gear (Figure 4e) is conical surface. Contact areas between these five new cycloidal gears and their corresponding needle teeth are large. Contact stress is uniform. They have high transmission efficiency and long service life.

a) Concave cycloidal gear   b) Drum shaped Cycloidal gear

c) Spherical cycloidal gear   d) Oblique cycloidal gear

e) Conical cycloidal gear
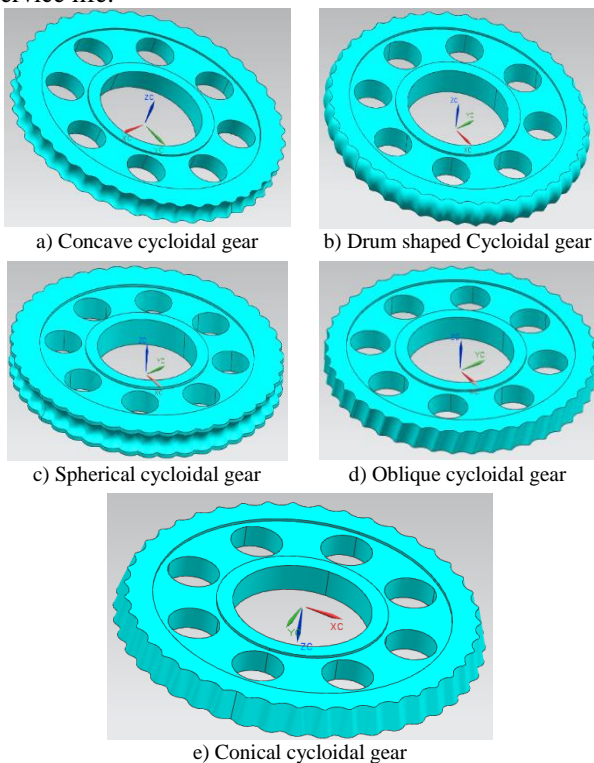
FIGURE 4 5 kinds of variable cross-section cycloid gear

4.2 MODAL ANALYSIS

If structural design of cycloidal gear is unreasonable, then the cycloid reducer at work will generate noise and vibration. It is necessary to conduct finite element modal analysis for the five kinds of new cycloid gears. Figure5 shows the comparison of natural frequency among the five kinds of variable cross-section cycloid gears and ordinary cycloid gear. As is shown, the first order natural frequency of ordinary cycloid gear is 2094Hz and the first order natural frequencies of the five new variable cross-section cycloid gears are 2068Hz, 2019Hz, 1884Hz, 2039Hz and 1963Hz, respectively. The first six natural frequencies of the 5 variable cross section cycloid gears are less than that of ordinary cycloid gear; so, it is easier to avoid the resonance for the 5 variable cross section cycloid gears. Changing trends of the first six natural frequencies of the 5 variable cross section cycloid gears are consistent with that of ordinary cycloid gear. Thus dynamics of the five new variable cross-section cycloid gears are consistent with that of ordinary cycloid gear .The front sixth modal shapes of three kinds of cycloid gear are shown in Figure 6,7 and 8 ,the shape of which is concave, drum-shaped and spherical respectively. The front sixth modal vibration shapes of these three kinds of cycloidal gear are very similar. The main deformation parts locate in the middle and edge of cycloidal gear. Deformation of spherical cycloid gear is bigger than that of concave and drum shaped cycloid gear.
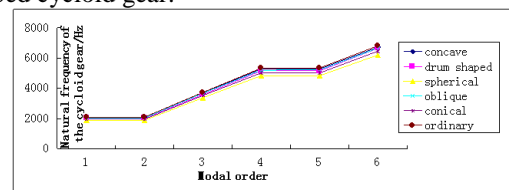
FIGURE 5 The comparison of natural frequency between variable cross-section cycloid gear and ordinary cycloid gear

a) The first order   b) The second order

c) The third order   d) The fourth order

e) The fifth order   f) The sixth order
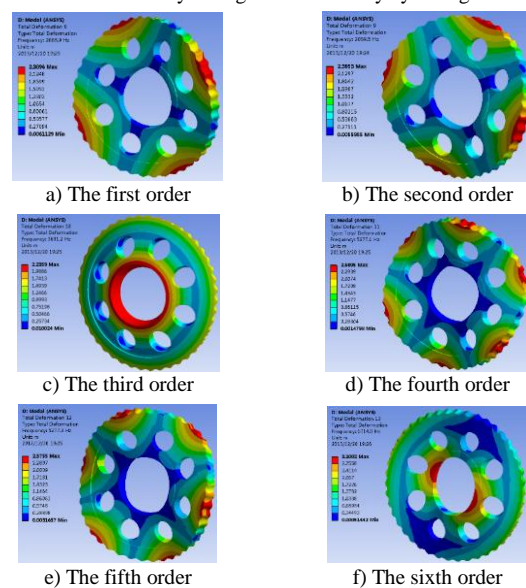
FIGURE 6 The modal vibration pattern map of concave cycloidal gear from the first to the sixth order

a) The first order

b) The second order

c) The third order

d) The fourth order

e) The fifth order
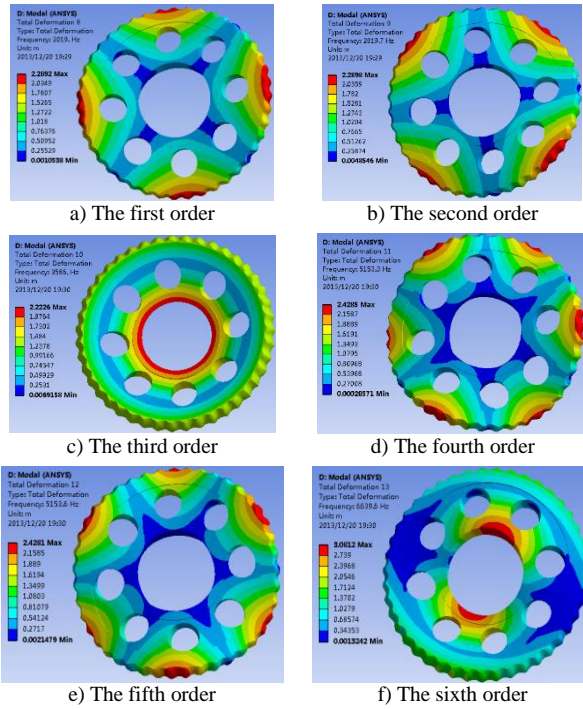
f) The sixth order

FIGURE 7 The modal vibration pattern map of drum shaped cycloidal gear from the first to the sixth order

a) The first order

b) The second order

c) The third order

d) The fourth order

e) The fifth order
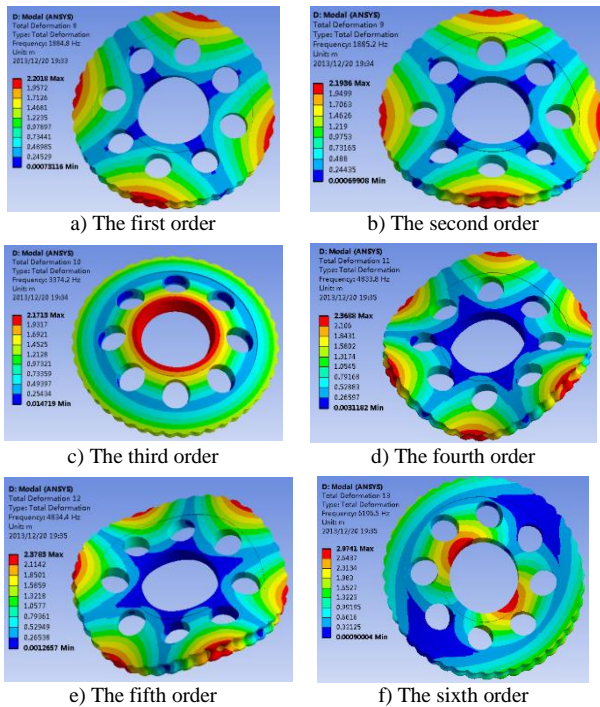
f) The sixth order

FIGURE 8 The modal vibration pattern map of spherical cycloidal gear from the first to the sixth order

5 Impact of material

Due to the development of high-performance engineering plastics, application of plastics in industry is accelerated. Materials used to produce gears are no longer limited to metals, plastics are used as well. Finite element model analysis of ordinary cycloid gear, with different materials,

including bearing steel, alloy steel and plastics, are performed. Material parameters of cycloid gear are shown in Table 1.

TABLE 1 Material parameters of cycloid gear

Material	Modulus of elasticity (MPa)	Poisson's ratio	Density (kg/m³)
Bearing Steel 20CrMnTi	2.07×10^5	0.25	7.8×10^3
alloy steel SNCM220	2.2×10^5	0.3	7.8×10^3
Engineering Plastics MC901	36	0.4	1.41×10^3

TABLE 2 The first six natural frequencies of ordinary cycloid gear

Modal order	Bearing Steel 20CrMnTi	alloy steel SNCM220	Engineering Plastics MC901
	Natural frequency (Hz)		
1	2094	2143	63
2	2094	2144	63
3	3734	3860	116
4	5335	5463	162
5	5337	5467	162
6	6810	6996	209

a) The first order

b) The second order

c) The third order

d) The fourth order

e) The fifth order

f) The sixth order
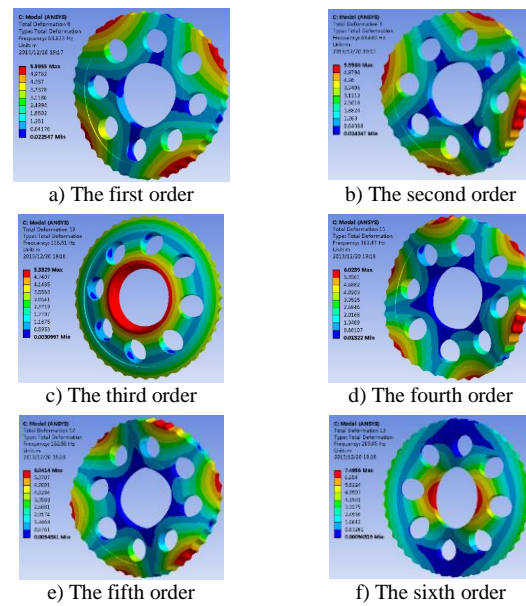
FIGURE 9 The first-sixth modal vibration pattern map of ordinary cycloid gear with plastic MC901

When free modal analysis is conducted the front six natural frequencies of ordinary cycloid gear are close to zero. It is the rigid modal. Research on rigid modal is of no real meaning. So each modal frequencies of the cycloid gear after the seventh modal are extracted. The front six modal nonzero modal that is after the seventh modal of cycloidal gear is researched and analysed in this article. The front six natural frequencies of cycloid gear with 3 kinds of different materials are shown in Table 2. As can be seen from Table 2, that the minimum natural frequency of the cycloidal gear with these three materials are 2094Hz, 2143Hz and 63Hz respectively. The natural frequency of cycloid gear with plastic MC901 is significantly less than that of cycloid gear with bearing steel or alloy steel. The

natural frequency of ordinary cycloid gear is gradually increasing with the modal order increases. The nonzero order front six vibration modal of the ordinary cycloid gear with plastic material MC901 is shown in Figure 9. As can be seen from the Figure 6 that the first-order and the second-order vibration mode of cycloidal gear are the circumference modes, the central part of cycloidal gear has almost not distortion. The third order vibration mode of cycloidal gear is an umbrella-type vibration and the deformation of the central part is the largest. The forth order and the fifth-order vibration mode of cycloid gear is torsional vibration mode, the central part of the cycloid gear has no distortion. The sixth vibration mode of cycloid gear is radial mode; the radial deformation of the central portion is the largest, peripheral part has not distortion.

6 Impact of the thickness

The thickness change of cycloid gear affects the dynamic characteristics of the cycloidal gear. To reveal this, modal analysis of ordinary cycloid gear with thicknesses of 11mm, 13mm, 15mm, 17mm and 19mm is conducted by using finite element software and the front sixth natural frequencies of cycloid gear with six kinds of different thicknesses are obtained. Details are shown in Table 3. From the data, the natural frequencies of cycloidal gear with different thickness have an increasing trend with the increase of modal order. Simultaneously the natural frequency of the same modal order is gradually increasing with the increase of cycloid gear thickness.

TABLE 3 The front six natural frequencies of cycloid gear with different thickness

Cycloidal gear thickness (mm)	The first order (Hz)	The second order (Hz)	The third order (Hz)	The fourth order (Hz)	The fifth order (Hz)	The sixth order (Hz)
11	2860	2878	2880	3492	3493	5699
13	3750	3750	3759	4412	4413	6939
15	4546	4548	4592	5251	5253	8081
17	5279	5281	5363	6025	6027	9137
19	5948	5949	6070	6735	6737	9638

7 Conclusions

1) 5 kinds of variable cross-section cycloid gears, including concave, drum-shaped, spherical, oblique and conical cycloid gears, are studied, and finite element modal analysis are conducted. The front six natural frequencies of the 5 variable cross section cycloid gears are smaller than that of ordinary cycloid gear, so it is easier to avoid resonance for the 5 variable cross section cycloid gears. Changing trends of the front six order natural frequencies of the 5 variable cross section cycloid gears are consistent with that of ordinary cycloid gear. Thus, dynamics of the five new variable cross-section cycloid gears are consistent with that of ordinary cycloid gear.

2) The natural frequency of ordinary plastic cycloid gear is significantly lower compared to those of the bearing steel and alloy steel cycloid gears. Order modal

frequencies of cycloidal gears with these three materials show an increasing trend with the increasing of modal order. Modal shape of cycloidal gear include circumferential mode, umbrella-type mode, Torsional mode and radial mode. The deformation of cycloid gears in the peripheral portion and the central portion is the greatest.

3) The natural frequencies of cycloidal gear with different thickness have an increasing trend with the increase of modal order. Simultaneously the natural frequency of the same modal order is gradually increasing with the increase of cycloid gear thickness.

Acknowledgments

References

[1] Yan H S, Lai T S 2002 Geometry design of an elementary planetary gear train with cylindrical tooth-profiles *Mechanism and machine theory* **37**(8) 757-67
[2] Lai T S 2006 Design and machining of the epicycloid planet gear of cycloid drives *The International Journal of Advanced Manufacturing Technology* **28**(7) 665-70
[3] Lai T S 2005 Geometric design of roller drives with cylindrical meshing elements *Mechanism and machine theory* **40**(1) 55-67
[4] Lai T S 2006 Geometric design of a pinion with two circularly arrayed conical teeth for roller drives *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science* **220**(9) 1405-12
[5] Lei L, Shi X C, Guan T M 2012 Finite Element Analysis for Cycloid Gear and pin Teeth of FA Cycloid Drive Based on ANSYS *Applied Mechanics and Materials* **215** 1197-1200

[6] Thube S V, Bobak T R 2012 Dynamic Analysis of a Cycloidal Gearbox Using Finite Element Method *AGMA Technical Paper* 12FTM18
[7] Zhang X Y, Dai X J 2012 Meshing Stiffness Analysis of Four Ring-Plate-Type Pin-Cycloidal Gear Planetary Drive *Applied Mechanics and Materials* **229** 499-502
[8] Zhang X Y, Dong S J 2012 The Input Axis' Mode Analysis of the Four Ring-Plate-Type Cycloid Gear Planetary Drive *Advanced Materials Research* **588** 238-41
[9] Biernacki K 2011 Methods of Increasing Load ability for the Plastic Cycloidal Gears *Key Engineering Materials* **1390**(490) 156
[10] Nam W K, Shin J W, Oh S H 2013 Design of thin plate-type speed reducers using balls for robots *Journal of Mechanical Science and Technology* **27**(2) 519-24
[11] Xiao J, He W 2009 Modal finite element analysis of RV reducer cycloid *Mechanical Engineers* **9** 46-47

[12] Shun, Jiangying Lan, Weidong 2013 Modal finite element analysis of cycloid reducer for wind turbine pitch drive *Dalian Jiaotong University* **34**(1) 44-8 *(in Chinese)*

[13] Qin Guang Yue, Zhu Hua, Chen Xiaohui, et al. 2011 Modal Analysis of shearer cycloid gear based on ANSYS *Coal Mine Machinery* **32**(2) 86-8 *(in Chinese)*

[14] Woody, Zhang Youchen 2009 Finite element analysis of multi tooth difference cycloid gear based on ANSYS *Mechanical design* **5** 40-2 *(in Chinese)*

Authors

Jianmin Xu, born in October, 1981, Hunan Province, China

Current position, grades: PHD candidate at Huaqiao University, China.
Scientific interest: digital design and manufacturing.
Publications number or main: 6 patents, more than 10 research papers.

Lizhi Gu, born in November, 1956, China

Current position, grades: PhD supervisor, professor of Huaqiao University, China.
Scientific interest: digital design and manufacturing, metal cutting and advanced manufacturing technology.
Publications number or main: 7 patents, more than 120 research papers.

Shanming Luo, born in October, 1969, China

Current position, grades: Ph.D., professor at Xiamen University of Technology, China.
Scientific interest: mechanical drive innovative design, precision plastic forming technology, CNC technology and applications.
Publications number or main: 4 patents, more than 110 research papers.

Stochastic resonance induced by over-damped fractional Langevin equation with α-stable noise

Yong-jun Zheng[1, 2*], Fei Wang[2], Shan-an Zhu[1]

[1]*College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China*

[2]*College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China*

Abstract

Stochastic resonance phenomenon induced in a system described by over-damped fractional Langevin equation with α-stable noise is investigated. When there is no external α-stable noise, the stochastic resonance is observed in case of the fractional order less than one certain threshold. By applying α-stable noise, the influences of the noise intensity and characteristic exponent of α-stable noise on the occurrence of stochastic resonance phenomenon are characterized. We find that the proper noise intensity enlarges the peak value of output power spectrum which is significant for stochastic resonance. Adjusting the noise intensity, the behaviour of signal-to-noise ratio is non-monotonic and with a maximum value. Under the same conditions, the lower value of characteristic exponent of α-stable noise leads to the smaller noise intensity to achieve stochastic resonance.

Keywords: stochastic resonance, over-damped fractional Langevin equation, α-stable noise

1 Introduction

The fractional calculus has a long history since it was first described by G.W. Leibniz [1], it allows the derivatives or integrals to be any non-integer order. Fractional calculus has been applied in many situations, such as viscoelasticity, confined geometries, biological tissues, thermoelasticity and control system etc. [2-6]. The classical derivatives and integrals form of fractional calculus are defined by Riemann-Liouville and Caputo.

Stochastic resonance (SR) has been widely investigated during past decades. It is a nonlinear phenomenon where a signal can be enhanced by adding noise. In the classical SR theory, models based on integer-order equation and double-well potential is defined to describe the resonance, which is characterized by the flow over the potential barrier [7]. In such a system, the occurrence of a single-well escape is a result of competition between damping and excitation. But this SR phenomenon is expected to be more complicated in system with memory effect, which can be introduced by hidden variables of non-viscous damping [8].

Recently, the phenomenon of stochastic resonance in the fractional order systems was investigated. The authors claimed that the stable steady states can be changed by fractional order damping and then lead to single- or double-well resonance behaviour [9]. SR was also investigated in the under-damped fractional Langevin equation, the signal-to-noise (SNR) and output signal's spectrum is found being non-monotonic, that indicates the SR phenomenon occur [10]. The authors also investigate that the SR phenomenon appears in the fractional order

system and this characteristic can be used to detect the weak signal [11]. Peng Hao etc. studied the SR in the over-damped bistable system with chirp signal, the results show that there is certain relation with the chirp signal frequency and SR [12].

In most of the previous studies of SR in fractional-order system, the noise was assumed to be white Gaussian noise. It is commonly used to describe various phenomena due to the Central Limit Theorem. White Gaussian noise is just an ideal case for fluctuations, however, in practical applications, non-Gaussian statistics is better to explain the additive noise, such as noise in communications channel and embedded wireless laptop transceivers [13], were found to be impulsive, so they cannot be characterized well using the Gaussian noise. The α-stable noise can describe the impulsive characteristic of noise much better, it can maintain the generation mechanism of natural noise and limit distribution of propagation conditions, match the actual data well. Gaussian noise is a particular example of it [14]. Its distribution follows heavy-tail stable law statistics with infinite variance; it can not only simulate the stable situation of noise, but also the impulsive status.

The interest to discuss α-stable noise in SR has just been started. The authors of [15] employ numerical methods to find the solution of stochastic Langevin equation and space fractional kinetic Equation, they studied the properties of the probability density function (PDF) of a bistable system driven by heavy tailed white symmetric Lévy noise. It is founded that in contrast to the bistable system driven by Gaussian noise, in the Lévy case, the positions of maxima of the stationary PDF do not

coincide with the positions of minima of the bistable potential. Tomasz Srokowski [16] discussed with generalized Langevin equation with double-well potential, the probability density distributions converge with time to a distribution similar to a Gaussian but tails have a power-law form. The SR phenomenon is emerged by means of spectral amplification.

In this paper, we consider the SR phenomenon induced by over-damped fractional Langevin equation with α-stable noise, in addition, by a period driving force. When the rate of the jumping between the potential wells due to the α-stable noise coincides with the frequency of the oscillatory force, the SR is observed. We discuss this phenomenon and demonstrate the influence to SR by various sets of the model parameters, signal-to-noise (SNR) and power spectrum amplification are taken as characteristic.

The paper is organized as follows. In section II, we introduce the models and methods of over-damped fractional Langevin equation and α-stable noise. The related potential function and density function of α-stable distribution are obtained in different cases. In section III, the variation of output signal and power spectrum with noise intensity is analysed, and we demonstrate how model parameters such as characteristic exponent modify its properties, in particular the SNR function. Section IV is some discussions and conclusions.

2 Models and methods

Consider an over-damped fractional Langevin equation [15, 16] driven by α-stable noise

$$ {}_0^C D^p x(t) + \frac{dV(x)}{dx} = F_1(t) + E\eta(t) , \qquad (1) $$

where ${}_0^C D^p x(t)$ is the p order fractional order derivative to $x(t)$ by using Caputo's definition, and $0 < p < 1$, the Caputo's definition is written as

$$ {}_a D_t^p f(t) = \frac{1}{\Gamma(n-p)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{p-n+1}} d\tau . \qquad (2) $$

We take a lower limit $a = 0$ for the above definitions. $F_1(t) = A\cos(2\pi\omega t)$ is an external signal with amplitude A and frequency ω, and $\eta(t)$ denotes the α-stable noise with characteristic exponent α ($\alpha \in (0,2]$), which obey to the α-stable distribution. E is the intensity of α-stable noise. When $\alpha = 2$, $\eta(t)$ becomes a Gaussian noise. The potential function $V(x)$ in the Equation (1) is defined as

$$ V(x) = -ax^2 / 2 + bx^4 / 4 \ (a > 0, b > 0) \qquad (3) $$

$V(x)$ is a symmetric double-well potential, as shown in Figure 1. There are two minima located at $\pm x_m$, they are separated by a potential barrier with height

$\Delta V = a^2 / (4b)$. Without the extern periodic forcing or the forcing is too weak, the particle cannot roll periodically from one potential well into the other on. From Equation (2) the Caputo's definition of fractional order differential, we can see the fractional order p relates with the memory characteristic, the bigger p means the memory characteristic much less, when $p \to 1$, it turns into the integer differential order, which indicates totally loss memory, while $p \to 0$, it differential equal to constant 1, that indicates the same memory characteristic to the speed of each time.
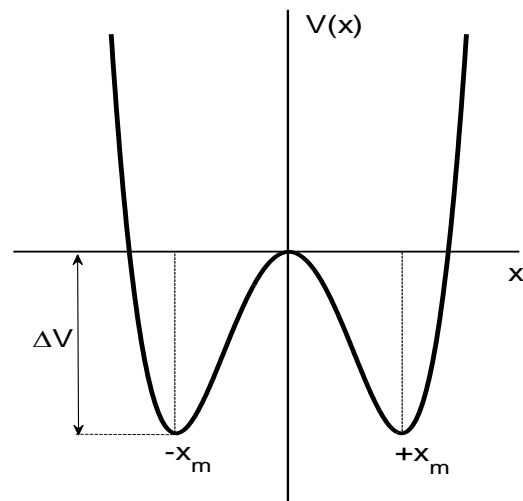
FIGURE 1 Sketch of the double-well potential V(x)

When there is no noise applied to the Equation (1), just in the presence of periodic driving force $A\cos(2\pi\omega t)$, by changing the fractional order p, the double-well potential is tilted back and forth, thereby the potential barriers of the right and the left well will be successively raised and lowered, respectively, in an anti-symmetric manner. Figure 2 shows the numerical simulation of Equation (1) without the noise $\eta(t)$, the other parameters are $a = b = 1$, $E = 0.3$, $f = 0.01$. Figure 2(a) indicates the curves when fractional order p changes from 0.9 to 0.1 with step 0.1; Figure 2(b) shows the curves when p changes from 0.3 to 0.2 with step 0.01. From the figures it can be concluded that as the fractional order p attenuated from (0,1). The particle is doing partial periodic motion around the balance point $x = 1$ or $x = -1$ with frequency $f = 0.01$, when p reaches a threshold p_t, the particles hop the potential barrier top which takes place at $x = 0$ into the other well, thus do the periodic motion between $x = \pm 1$ with the centre at $x = 0$. From the simulation results, the threshold p_t is 0.29. When p is less than p_t, the particles can hop the potential barrier top without external noise energy, thus the stochastic resonance phenomenon cannot be appeared. While p is greater than p_t, the particles just do partial periodic motion around one well, it only needs the synchronized action with the external noise to produce stochastic resonance phenomenon.

a)   b)
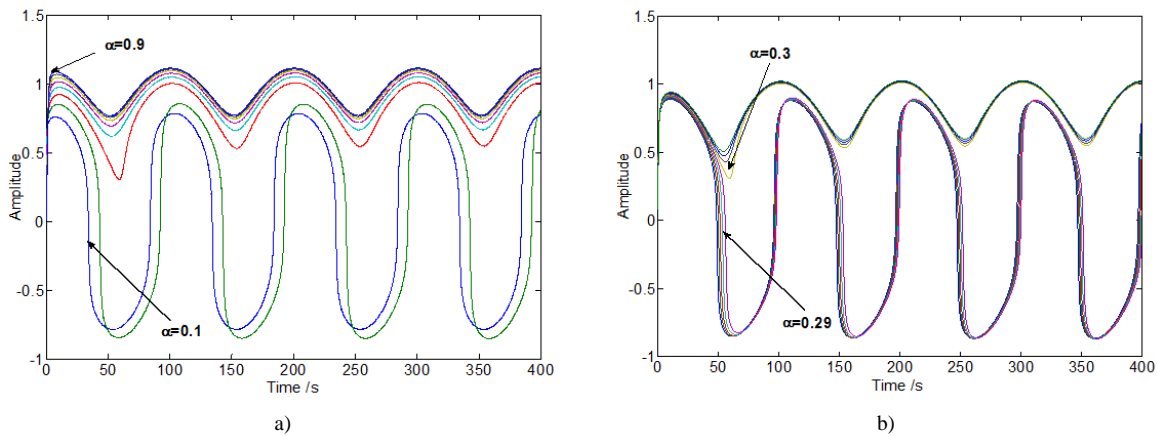
FIGURE 2 Time domain of output signal with different fractional order *p*

α-stable noise $\eta(t)$ obeys α-stable distribution, whose characteristic function is [14]:

$$\varphi(t) = \begin{cases} \exp(-\gamma^{\alpha} |u|^{\alpha} [1 - i\beta(\tan\frac{\pi\alpha}{2})\text{sign}(u)] + i\delta\mu), \alpha \neq 1 \\ \exp(-\gamma |u| [1 + i\beta\frac{2}{\pi}\log|u|\,\text{sign}(u)] + i\delta\mu), \alpha = 1 \end{cases},(4)$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma \geq 0$ and $\delta \in \mathbb{R}$, sign(u) function is -1 for a negative number, 0 for the number zero, or +1 for a position number. α-stable noise characteristic function is determined by four parameters: characteristic exponent α, scaling parameter γ, symmetry parameter β and location parameter δ. A small value of α will imply considerable probability mass in the tails of the distribution. It corresponds to the Gaussian distribution (for any β) when $\alpha = 2$, a Cauchy distribution with $\alpha = 1, \beta = 0$ and a Lévy distribution with $\alpha = 1 / 2, \beta = 1$.

Figure 3 illustrates the PDF of α-stable noise with different parameters. Figure 3(a) shows the relation of symmetric α-stable noise PDF with different characteristic exponent parameters α, Figure 3(b) displays the relation of skewed α-stable PDF with different symmetry parameters β.

In this paper the fractional order operator is approximated by a refined Oustaloup recursive filter [17] in a specified frequency range (ω_b, ω_h) and of order *N*. It is given by

$$s^{\alpha} \approx \left(\frac{d\omega_h}{b}\right)^{\alpha} \left(\frac{ds^2 + b\omega_h s}{d(1-\alpha)s^2 + b\omega_h s + d\alpha}\right) G_p, \qquad (5)$$

where G_p, ω_k, ω_k' can be computed from

$$G_p = \prod_{k=-N}^{N} \frac{s + \omega_k'}{s + \omega_k}, \ \omega_k = (b\omega_h/d)^{\frac{\alpha+2k}{2N+1}}, \ \omega_k' = (b\omega_b/d)^{\frac{\alpha-2k}{2N+1}}.$$

A good approximation is obtained with *b*=10, *d*=9.

3 Numerical results

We fix the parameters in Equation (1) as $a = b = 1$, $A = 0.3$, $\omega = 0.01$, $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 1.0$, $\delta = 0$ for α-stable noise $\eta(t)$, Equation (1) turns into:

$$\frac{d^p x(t)}{dt} = x - x^3 + 0.3\cos(2\pi \times 0.01) + E\eta(t) \qquad (6)$$

where *E* is the noise intensity.

3.1 THE VARIATION OF OUTPUT SIGNAL AND POWER SPECTRUM WITH NOISE INTENSITY

First we investigate the effects of α-stable noise on the evolution of *x*(*t*) under fractional order *p*=0.75. We choose the noise intensity *E*=1.5, The time domain and frequency spectrum of input signal and output signal *x*(*t*) are shown in Figure 3. Figure 4(a) shows the time domain of input signal which is the external periodic signal with α-stable noise, some sharp spikes are visible for the heavy tails of α-stable noise, Figure 4(c) illustrates the power spectrum of input signal, the peak amplitude of power spectrum is 0.1732 at frequency 0.01. Figure 4(b) shows the time domain of output signal, Figure 4(d) illustrates the power spectrum of output signal, the peak amplitude of power spectrum is 0.5198 at frequency 0.01, greater than 0.1732, which shows the SR significantly occurred. From Section II, we found that if there is no noise applied to the system, when fractional order *p* is greater than p_t, no SR phenomenon happened. From Figure 4 we know that the α-stable noise with proper intensity can cause the hopping of the particle between two potential wells, thus lead to the SR effect.

42
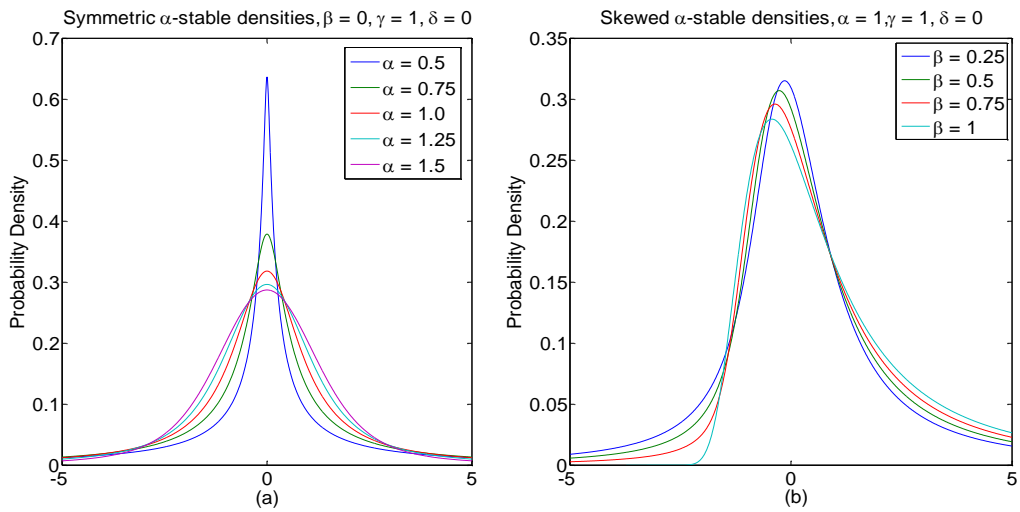
FIGURE 3 α-stable probability density functions
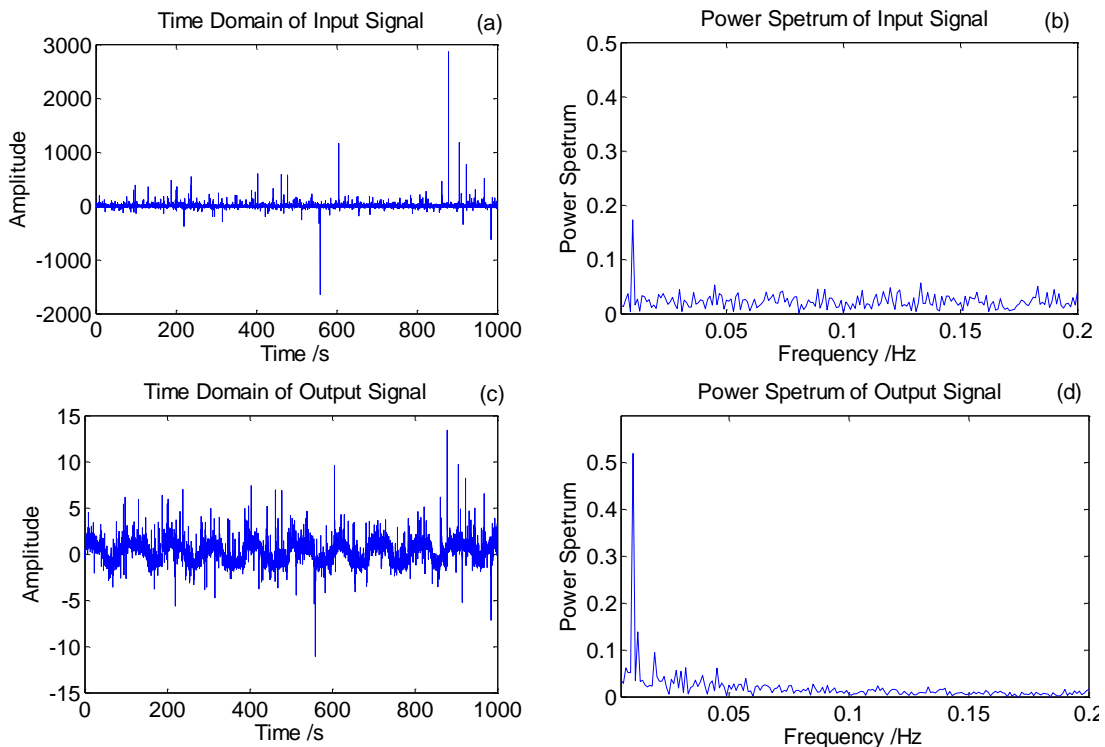
FIGURE 4 The time domain and power spectrum of input and output signal when intensity E=1.5

Figure 5 illustrates the statistics of particle oscillating back and forth between wells, it can be seen that the SR phenomenon is the enhancement of output signal via tuning the noise intensity, but when the noise intensity is bigger enough, the effect of SR gradually diminish. At a lower noise level, the particles oscillates at the minima of the potential wells for a long time and rarely switches between two potential wells, thus the periodic particles can hardly be visible at the other potential well. Under this circumstances, the periodic component of the output signal x(t) is primarily doing motion around the potential minima, which is interval motion in stochastic resonance, it is illustrated by Figure 5(a) where noise intensity E=0.7. However, when the noise intensity is increased to a certain value, the input-output synchronization effect happens, the periodic particles doing motion between two potential wells, we call it the interval motion in stochastic resonance. Figure 5(b) illustrated this phenomenon. Figure 5(c) shows that the synchronization vanishes when noise intensity is larger enough, that means the system flips too many times between its stable states within each forcing period, thus statistically irrelevant.

43
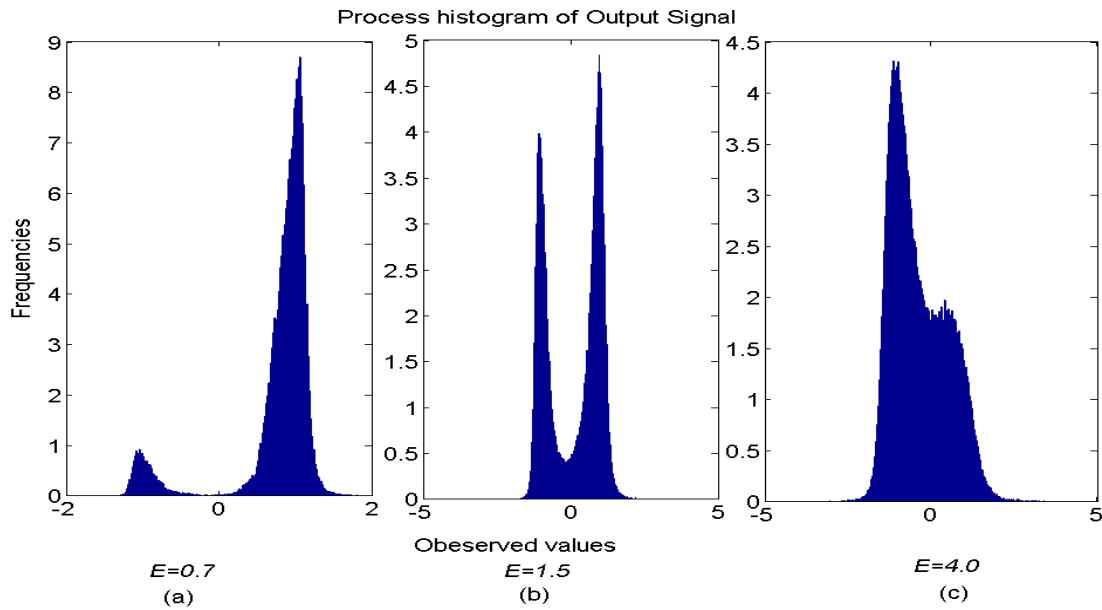
FIGURE 5 The process histogram of output signal

3.2 THE SNR FUNCTIONS WITH DIFFERENT CHARACTERISTIC EXPONENT

SNR is often taken as one quantitative indicator to demonstrate the SR phenomenon [18]. The definition is:

$$SNR = \frac{1}{S_N(\Omega)} \lim_{\Delta\omega \to \infty} \int_{\Omega-\Delta\omega}^{\Omega+\Delta\omega} S(\omega)d\omega , \qquad (7)$$

where $\int_{\Omega-\Delta\omega}^{\Omega+\Delta\omega} S(\omega)d\omega$ represents the power carried by the signal, $S_N(\Omega)$ represents the noise power spectrum near the frequency, and $S(\omega)$ denotes the power spectrum of signal.

Figure 6 shows the SNR versus the noise intensity E with different characteristic exponent of α-stable noise, the other parameters being kept same. The values of SNR decrease with the noise intensity E at first, then begin to increase, and when the noise intensity reach to a critical value E_{SR}, the values of SNR achieve a maximum and after that decrease again. Under the different characteristic exponent α, the SNR is clearly non-monotonic, thus indicates the occurrence of SR phenomenon. As increasing the characteristic exponent α the SNR shifts towards bigger values of noise intensity.

4 Discussions and conclusions

In this paper, the properties of over-damped fractional Langevin equation with α-stable noise have been studied. In case of no external α-stable noise, the stochastic resonance phenomenon is observed when fractional order is less than one certain threshold. When fractional order is

greater than the certain threshold, the SR is not appeared. However, by applying the α-stable noise, even at the situation with larger fractional order, the SR phenomenon is occurred, by comparing with the output power spectrum of input signal and output signal, we investigate that the proper noise intensity enhance the peak value of output power spectrum, the behaviour of SNR is non-monotonic, there is a maximum value when the noise intensity changes, thus is the typical SR phenomenon. We also find that at the same conditions, the smaller of the characteristic exponent of α-stable noise, the lower of noise intensity to achieve the SR.
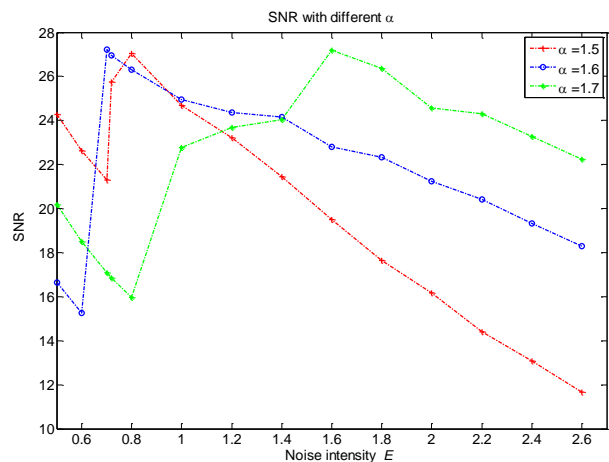
FIGURE 6 SNR versus noise intensity E with different α

Acknowledgments

References

[1] Machado J T, Kiryakova V, Mainardi F 2011 *Communications in Nonlinear Science and Numerical Simulation* **16** 1140-53

[2] Meral F C, Royston T J, Magin R 2010 *Communications in Nonlinear Science and Numerical Simulation* **15** 939-45

[3] Jeon J H, Metzler R 2010 *Phys Rev E* **81** 021103

[4] Magin R L 2010 *Comput Math Appl* **59** 1586-93

[5] Sherief H H, El-Sayed A M A, El-Latief A M A 2010 *Int J Solids Struct* **47** 269-75

[6] Mozyrska D 2014 Discrete *Dyn. Nat Soc* 183782

[7] Gammaitoni L, Hanggi P, Jung P, Marchesoni F 2009 *Eur Phys J B* **69**(1)

[8] Ruzziconi L, Litak G, Lenci S 2011 *Journal of Vibroengineering* **13** 22-38

[9] Goychuk I, Kharchenko V 2012 *Phys Rev E* **85** 051131

[10] Zhong S, Wei K, Gao S, Ma H 2013 *J Stat Phys* **1**

[11] He G T, Luo M K 2012 *Chinese Phys Lett* **29** 060204

[12] Peng H, Zhong S C, Tu Z, Ma H 2013 *Acta Phys Sin-ch* **62** 080501

[13] Nassar M, Gulati K, DeYoung M R, Evans B L, Tinsley K R 2011 *Journal of Signal Processing Systems for Signal Image and Video Technology* **63** 1

[14] Al-Talibi H 2013 Brazilian *Journal of Probability and Statistics* **27** 544-52

[15] Sliusarenko O Y, Surkov D A, Gonchar V Y, Chechkin A V 2013 *European Physical Journal* **216** 133

[16] Srokowski T 2013 Eur *Phys* **86** 239

[17] Tricaud C, Chen Y Q 2010 *Comput Math Appl* **59** 1644-55

[18] Gao S L, Zhong S C, Wei K, Ma H 2012 *Acta Phys Sin-ch* **61** 100502 (*in Chinese*)

Authors

Yongjun Zheng, 1977-02-10, Ninghai, Zhejiang, China.

Current position, grades: Ph.D candidate, associate professor of College of Metrology and Measurement Engineering at China Jiliang University.
University studies: Zhejiang University, China.
Scientific interest: signal processing, fractional system, intelligent measurement.
Publications: 12 papres.

Fei Wang, 1989-07-04,Gong-an, Hubei, China.

Current position, grades: master student.
University studies: China Jiliang University.
Scientific interest: signal processing, fractional system, intelligent measurement.
Publications: 1 paper.

The simulation and analysis of fish school behaviours with different body lengths

Huanda Lu[1, 3], Xin Yu[3*], Ying Liu[2], Kangsheng Liu[1]

[1]*Department of Mathematics, Zhejiang University, Hangzhou, 310027, China*

[2]*Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071, China*

[3]*Ningbo Institute of Technology, Zhejiang University, Ningbo, 315100, China*

Abstract

In this paper, through simulating and modelling of fish school with different body length, we study the influences of the body length difference of fish school on spatial structure and group behaviour. Based on attraction/repulsion model, we obtain three typical spatial structure of group with different model parameters: mixture structure, periphery structure and front-back structure. Moreover, we analyse the polarization index and average angular speed of group with different model parameters and get the corresponding relationship between these indices and model parameters. The results obtained in this paper coincide with the phenomenon observed in the natural world and the methods provide an effective way to study the fish school behaviour.

Keywords: computer simulation, attraction/repulsion model, aggregation behaviour, spatial structure

1 Introduction

Many biology systems are self-organizing in nature, such as fish school, bird flocks, grasshopper, ants etc. The aggregation behaviours in animals not only help protect themselves against predators, but also facilitate mate choice and access to information (such as the location of food sources or predators, migratory routes, etc.). Schooling of fish as a common aggregation phenomenon in nature is attracting a lot of biology, cognitive science, psychology, computer science, physics and other related fields of scientists to attempt to describe, explain and predict the structure and behaviour of aggregation. However, it is very difficult to study the fish schooling in the wild. Recently, computer simulation technology has become important tools in fish school behaviour research. Many models of aggregation offer researchers a way to investigate how the interplay of individual behaviour makes aggregation possible and leads to different aggregate-level behaviours [1-7].

Many fish school behavioural models stand on Breder's model [8, 9], Breder considered the individual in the group as atoms in the crystal and described the mechanical model using atomic attraction and repulsion analogue. Inspired by Breder's work, many Attraction/Repulsion (AR) models and their variants emerged. AR models consider the individuals in the group as independent agent, the agent attempt to maintain a minimum distance (within zone of repulsion) between themselves and others at all times, and they tend to be attracted towards other individuals and to align themselves with neighbours [1, 6]. For example, Couzin et al. [1]

modelled the collective behaviour using three simple rules (repulsion, attraction, alignment) and four common collective behaviour patterns are produced in the model. Jennifer et al. [7] extended the AR model from discrete to continuous. Vincent [10] presented numerical simulations of an animal grouping model based on individual behaviours of attraction, alignment and repulsion, and investigated how some factors such as the number of individuals, the number of influential neighbours and the strength of the alignment behaviour impacting on internal spatial structures. Rune et al. [11] studied how to adjust the parameters of the herring spawning behaviour model, resulting in some particular spatial structure. Many studies have been verified that the simulation results of AR model are consistent with field experiments in many fish and other animal aggregations, e.g. Starling [12], mosquito fish [13], surf scoters [11], golden shiners [14].

All the above models assumed that fish schools composed of individuals with the same body length, and thus their interaction zones are the same as well. But a large number of observations suggested that there are individual differences exist in fish school [15, 16]. In addition, fisheries often capture the uniform body length fishes at the beginning of fishing season, but get the varying body length individuals in late. This indicates that the fish school composed of diverse body length individuals is a common phenomenon in nature. However, the spatial structure and behaviour of such diverse collective systems are unclear.

In this paper, we consider the body length of individuals in fish school as a continuous random variable,

and simulate the fish school behaviours with different body lengths. The main objective are to:

1) to develop a mathematics model for fish school which is composed of different body length individuals.

2) to analyse the spatial structure and aggregation behaviour generated by the model.

3) to reveal the transition rules of spatial structure and behaviour when the model parameters are changed.

2 The Model

We assume that the *N* number of individuals with different body length in continuous three-dimensional space. Time is partitioned into discrete time steps with a regular spacing τ. In each time step *t*, the *i-th* individual with body length BL_i access the position $r_i(t)$ with direction vector $d_i(t)$. The individual's length is assumed to obey the Gaussian distribution with mean *m* and standard deviation σ. Interaction between individuals is restricted to the directional component, and the speed of individuals is configured to constant *S*.

In this research, we adopt a biological model concept by Couzin [1], which is based on the observational and empirical investigation of interaction of animal behaviour with its neighbors in the aggregation phenomenon. We divide the individual's perception range into three zones: zone of repulsion (ZOR), zone of attraction (ZOA) and zone of orientation (ZOO). These zones are spherical, except for a volume behind the individual within which neighbours are undetectable. This "blind area" is defined as a cone with interior angle $(360-\varphi)$, where φ is defined as the field of perception (see, Figure 1) [1]. Unlike fixed the perception field for all individuals in the Couzin's model, we assume that the individual perception range related to their body length. Let:

$$r_{ri} = \alpha \times BL_i, \tag{1}$$

$$r_{oi} = \beta \times BL_i, \tag{2}$$

$$r_{ai} = \gamma \times BL_i, \tag{3}$$

where r_{ri}, r_{oi}, r_{ai} represent the radius of three zones, and, α, β, γ are the perception coefficient of each zones. The interaction rules of individuals in our model extended from refs. [1, 8, 9]:

1) If n_r neighbours are present in the zone of repulsion at time *t*, individual *i* responds by moving away from neighbours within this zone:

$$d_{ri}(t+\tau) = -\sum_{j=1, j\neq i}^{n_r} \frac{r_{ij}(t)}{|r_{ij}(t)|}, \tag{4}$$

where $r_{ij} = (r_j - r_i)/|r_j - r_i|$ is the unit vector in the direction of neighbour *j*. This rule has the highest priority in the model, so the desired direction of next time is:

$$d_i(t+\tau) = \frac{d_{ri}(t+\tau)}{|d_{ri}(t+\tau)|}. \tag{5}$$

2) If no neighbours are within the zone of repulsion, the individual responds to other within the zone of orientation and the zone of attraction. An individual will attempt to align itself with neighbours within the zone of orientation, giving the direction:

$$d_{oi}(t+\tau) = \sum_{j=1}^{n_o} \frac{d_j(t)}{|d_j(t)|}, \tag{6}$$

and towards the positions of individuals within the zone of attraction

$$d_{ai}(t+\tau) = \sum_{j=1}^{n_a} \frac{r_{ij}(t)}{|r_{ij}(t)|}, \tag{7}$$

where n_o and n_a are the individuals' number in the zone of orientation and attraction. If neighbours are only found in the zone of orientation, then the desired direction of next time is

$$d_i(t+\tau) = \frac{d_{oi}(t+\tau)}{|d_{oi}(t+\tau)|}. \tag{8}$$

If neighbours are only found in the zone of attraction, then the desired direction of next time is

$$d_i(t+\tau) = \frac{d_{ai}(t+\tau)}{|d_{ai}(t+\tau)|}. \tag{9}$$

If neighbours are found in both zones, then the desired direction of next time is

$$d_i(t+\tau) = \frac{(d_{oi}(t+\tau) + d_{ai}(t+\tau))}{2}. \tag{10}$$

After the above process has been performed for every individual they turn towards the direction vector $d_i(t+\tau)$ by maximum turning rate θ. Provided the angle between $d_i(t)$ and $d_i(t+\tau)$ is less than the maximum turning angle $\theta\tau$, then $d_i(t+\tau) = d_i(t)$. In our model, the initial position and initial direction are generated by random and the next position of individual are computed by:

$$r_i(t+\tau) = r_i(t) + d_i(t+\tau) \times \tau \times s + \varepsilon, \tag{11}$$

where ε is the random disturbance taken from a spherically wrapped Gaussian distribution with standard deviation.
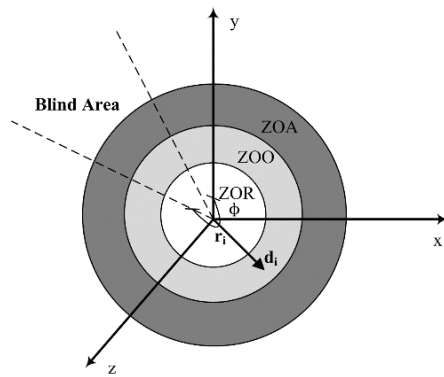
FIGURE 1 Diagram of three zones of interaction and the blind area associated with individual *i*

3 Simulation Experiments

According to previous model and the parameters showed in Table 1, the experiments were simulated through fixed the perception coefficient of ZOR *α* and changed the perception coefficient of ZOO *β*, the perception coefficient of ZOA *γ* and the standard deviation *σ*. Each experiment was repeated 5 times with 1000 time steps. The following analyses describe the influence in spatial structure and aggregation behaviour as the body length changed.

TABLE 1 Model parameters

Symbol	Parameter	Values
N	Number of individuals	256
$α$	Radius coefficient of ZOR (Unit[a])	1
$β$	Radius coefficient of ZOO (Unit[a])	2-10
$γ$	Radius coefficient of ZOA (Unit[a])	$4β$
m	The mean body length of individuals (Unit[a])	10
$σ$	The standard deviation of individuals (Unit[a])	0-5
$φ$	Perception range (Deg.)	330
$τ$	Time step(Sec.)	0.1
$θ$	Maximum turn rate (Deg./Sec.)	440
s	Speed (Unit[a]/Sec.)	2

[a] Similar to ref. [1], the "unit" relates to the non-dimensionality of certain parameters in the model with the characteristic length scale being associated with the body length of fishes, and the rest of the model parameters can be scaled appropriately.

4 Spatial Structures

As the standard deviation of body length changed, the aggregation behaviour of the system exhibits sharp transitions between three aggregation spatial structures (Figure 2), which we have labelled as follows:

Mixture structure: different sizes of fish appear randomly anywhere in aggregation. This occurs when individuals have little body length standard deviation (Figure 2a).

Periphery structure: the individuals with greater body length distribute in the group centre, but the individuals with smaller body length aggregate around the centre. This occurs when individuals have middle level body length standard deviation (Figure 2b).

Front-back structure: with further increase the standard deviation of body length, some individuals with larger body size aggregate in the front and back of fish school (Figure 2c). This can be understood as the larger individual in the group to play a "leadership" role.
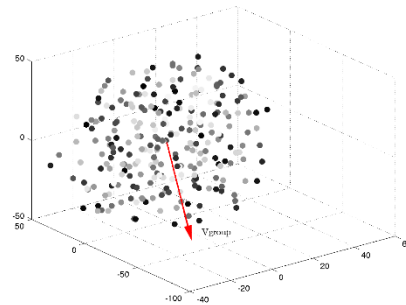
(a)
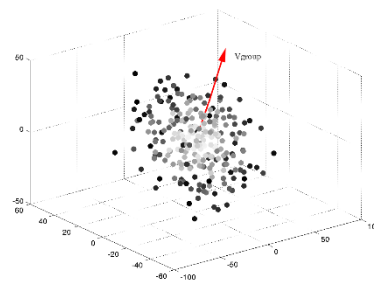
(b)
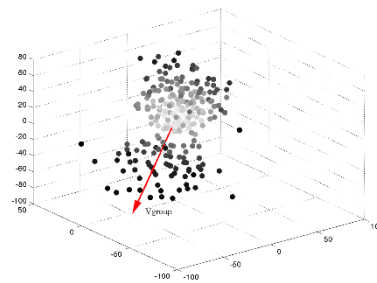
(c)

FIGURE 2 The aggregation spatial structure obtained by simulating (a) mixture structure (b) periphery structure (c) front-back structure. Small dots indicate individuals, each dot assign different level of gray color according to the fish's body length, the darker color indicate greater individual, the lighter color represents smaller individual, and the arrow indicates the direction of movement of fish school.

Figure 3 illustrates the contour of individual's body length after projected the spatial structure (Figure 2) to x-y plane. As shown in the Figure 3, the contour of mixture structure contains each other (Figure 3a), but the contour of periphery structure exhibits the obvious hierarchy structure (Figure 3b). In the contour of front-back structure, we can see that the contour which represents the largest body length only distribute in front of the most peripheral position of fish school (Figure 3c). In the y-z and x-z projection plane, similar results can be obtained.

(a)

(b)

(c)
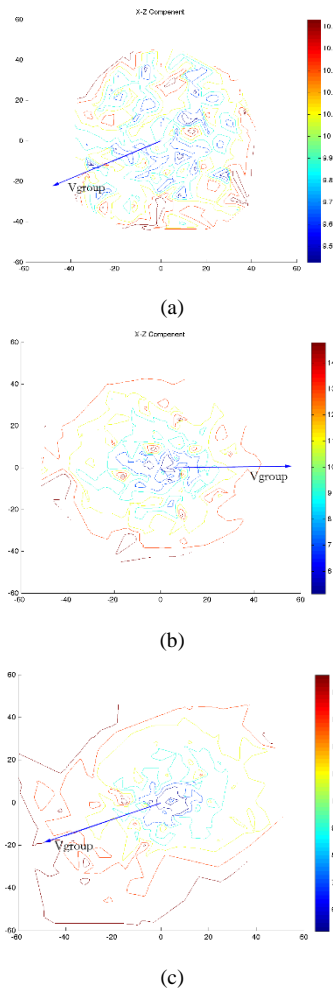
FIGURE 3 The contour of individual's body length projected to x-y plane (a) the contour of mixture structure (b) the contour of periphery structure (c) the contour of front-back structure. The arrow indicates the direction of movement of fish school, and different color represents the contour of different body length.

In order to further characterize the spatial structure of fish school with different body length standard deviation, we introduce the concepts of Average Center Distance of Group and p-Quintile Center Distance Ratio. Let the body length of individuals as continuous random variable X, the probability density function $p(x)$, X obey the Gaussian distribution $N(m, \sigma^2)$, where m and σ are the parameters that have showed in the Table 1. Let $0 < p < 1$, m_p is the p-Quintile of random variable X, then the below equation should be satisfied:

$$P(X > m_p) = p. \tag{12}$$

We define N_p as the set of individuals which body length is greater than or equal(s) m_p and:

$$\bar{r}(t) = \frac{1}{N} \sum_{i=1}^{N} r(t), \tag{13}$$

then, the Average Center Distance of Group and p-Quintile Center Distance Ratio can be written as

$$c(t) = \frac{1}{N} \sum_{i=1}^{N} \left\| \bar{r}(t) - r_i(t) \right\|, \tag{14}$$

$$c_p(t) = \frac{1}{c(t)} \frac{1}{|N_p|} \sum_{i \in N_p} \left\| \bar{r}(t) - r_i(t) \right\|, \tag{15}$$

where $|N_p|$ is the number of elements in the N_p. The Average Center Distance of Group represents the average distance from individual to group centre. The p-Quintile Center Distance Ratio indicates the ratio of individual's with greater body length to the average centre distance.
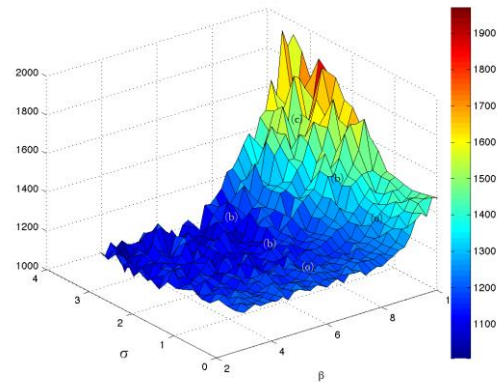
FIGURE 4 The average centre distance of group
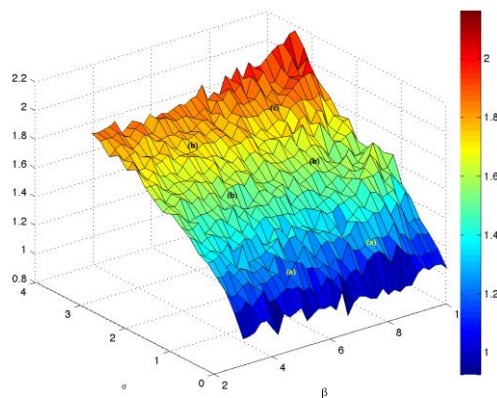
FIGURE 5 The centre distance ratio of p-Quintile

We can see from Figure 4, when the σ and β are smaller, the Average Center Distance of Group is also smaller. This indicates that the group has a higher degree of aggregation in this situation. When the σ and β increasing, the Average Center Distance of Group increases significantly. This shows that the group is more scattered. As shown in Figure 5, the value of p-Quintile Center Distance Ratio becomes larger with the σ and β increasing. The individuals that have bigger body length tend to aggregate in the peripheral position. Obviously, the variation of p-Quintile Center Distance Ratio is more influenced by σ compared with β. The markers (a-c) labelled in Figure 4 and Figure 5 correspond to the three spatial structures respectively.

5 Collective behaviour analyses

5.1 POLARIZATION INDEX

The polarization index [1] is defined as:

$$p_{group}(t) = \frac{1}{N}\left|\sum_{i=1}^{N} d_i(t)\right|. \tag{16}$$

The polarization index ($0 \le p_{group} \le 1$) characterizes the consistent level of individuals in the group. Figure 6 shows that polarization index exhibits lower values as the β and σ in lower level, and the fish school is poor consistency. As the β and σ increasing, the polarization index is also increase, and the group motion tend to be consistent. Especially, when $8.8 \le \beta \le 10$ and σ in lower level, the polarization index equals to 1, and the motion direction are consistent completely.
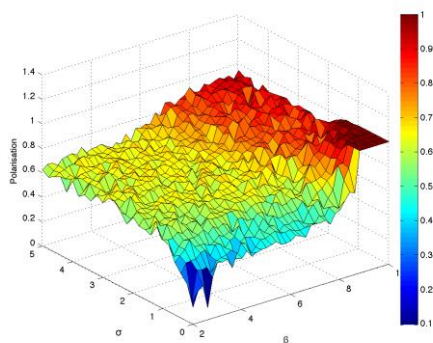
FIGURE 6 The polarization index of group
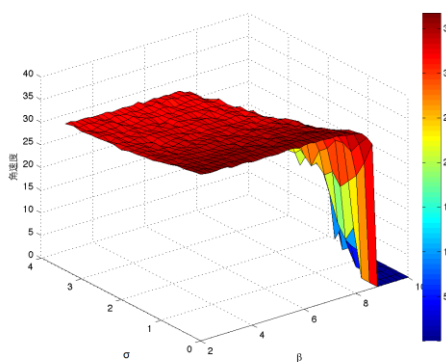
FIGURE 7 The average angular speed of group

5.2 AVERAGE ANGULAR SPEED

Let:

$$\omega_i(t) = \angle(v_i(t+\tau), v_i(t))/\tau , \tag{17}$$

represents the angular speed of individual i at time t, and $\angle(v_i(t+\tau), v_i(t))$ is the angular separation of two velocity direction. The average angular speed is defined as

$$\omega_{group}(t) = \frac{1}{N}\sum_{i=1}^{N} \omega_i(t). \tag{18}$$

As shown in the Figure 7, when $\beta < 8.8$, the average angular speed did not changed significantly as the σ changed, and the average angular speed is about 310-340 Deg./Sec.; when $8.8 \le \beta \le 10$, the average angular speed increase as the σ increasing. Especially, when the σ is in lower level, the average angular speed equals to 0, and this leads to the same conclusion with polarization index (the individuals tend to move in the same direction).

6 Conclusions

In this paper, we present a self-organizing fish school model with different body length, and use it to investigate how the body length difference (combined with other model parameters) affects the spatial structure and behavior. The results are consisting with many observational investigations in nature [17, 18]. The model can explain the spatial structure and dynamical behavior occurred in some fish school, and we also hope that our results may inspire empirical scientists to study spatial structure in schools of real fish.

This approach can be extended in future by:

1) adding other features that can lead to spatial structure or behavior changing, such as sex, hunger, disease etc.

2) inferring the model parameters from the real fish school data.

References

[1] Couzin I D, Krause J, James R, Ruxton G D and Franks N R 2002 *J Theor Biol* **218** 1-11
[2] Strombom D 2011 *J Theo Biol* **283** 145-51
[3] Giardina I 2008 *HFSP J* **2** 205-19
[4] Huth A, Wissel C 1992 *J Theor Biol* **156** 365-85
[5] Huth A, Wissel C 1993 *Comments Theor Bio* **3** 169-201
[6] Huth A, Wissel C 1994 *J Theor Biol* **75/76** 135-45
[7] Schellinck J, White T 2011 *Ecol Modelling* **222** 1897-911
[8] Jennifer M M, Allison K, Joao P, Juchem N, Louis F R 2012 *Bull Math Biol* **74** 536-61
[9] Breder C M 1951 *Bulletin of American Museum of Natural History* **98** 5-27
[10] Breder C M 1954 *Ecology* **35** 361-70
[11] Vincent M, Pierre A, Christophe L 2007 *Ecol Modelling* **201** 468-76
[12] Rune V, Georg S 2008 *Ecol Modelling* **214** 125-40
[13] Ballerini M, Cabibbo N, Candelier R, Cavagna A, Cisbani E, Giardina I, Lecomte V, Orlandi A, Parisi G, Procaccini A, Viale M, Zdravkovic V 2008 *PNAS* **105** 1-16
[14] James E H, Andrea P, Richard P M, Timothy M S 2011 *PNAS* **15** 18726-31
[15] Katz Y, Tunstrom K, Ioannou C C, Huepe C and Couzin I D 2011 *PNAS* **108**(46) 18720-5
[16] Brawn V M 1961 *Behaviour* **18** 177-198
[17] Kimura K 1934 *Bull Jap Soc Sci Fish* **3** 87-92

[18] Romey W L 1997 in *Animal groups in three dimensions* ed Parrish J K and Hamner W M Cambridge: Cambridge University Press 174-93

[19] Krause J 1993 *Oecologia* **93** 356-9

Authors

Huanda Lu, born in March, 1979, Zhejiang, China

Current position, grades: PhD candidate of Mathematics in Zhejiang University, China. Lecturer at Ningbo institute of Technology, Zhejiang University, China.
University studies: Master's degree in engineering at Zhejiang University, China, 2004.
Scientific interest: machine learning and its application in aquaculture, computer vision.
Publications: 10.

Xin Yu, born in June, 1977, Hebei, China

Current position, grades: associate professor, Ningbo institute of Technology, Zhejiang University, China.
University studies: Ph.D. degree in operation research and control theory at the Department of Mathematics in China.
Scientific interest: modeling and simulation of complex systems, data analysis.
Publications: 40.

Yin Liu, born in April, 1969, Shanxi, China

Current position, grades: professor, Key Laboratory of Marine Ecology & Environmental Sciences, IOCAS, China.
University studies: Ph.D. degree in Biosystem Engineering and Food Sciences from Zhejiang University in China.
Scientific interest: aquaculture engineering & ecology.
Publications: 47.

Kangsheng Liu, born in March, 1963, Sichuan, China

Current position, grades: doctor of mathematics, professor and doctoral supervisor in Zhejiang University, China.
University studies: Ph.D. degree at Fudan University in China.
Scientific interest: modeling and simulation of complex systems, data analysis, mathematical control and related fields.
Publications: 45.

Research on the software trustworthiness extended measurement based on IMC

Feng He[1], Haican Peng[2*], Kun Yao[3]

[1]*College of Mathematics, Physics and Information Engineering, Jiaxing University, China*

[2]*Computer Science and Engineering, Beifang University of Nationalities, China*

[3]*Hanzhong Branch Company of Shanxi Province of China Telecom China*

Abstract

This paper introduces two methods of extension measure based on model checking algorithm of interactive Markov chains (IMC) to decide the software trustworthiness. The first extended measurement is to establish multiple corresponding temporal logic relations for each software trustworthy attribute that affecting software trustworthiness, also is to use multiple temporal logic to describe a software trustworthy attribute, which is aim to measure the software trustworthiness on the multi-level and fine-grained. Then the paper will determine the measurement ultimately. The second extended measurement is to locate for the untrusted states, then find out the detail path and detail parameters of the path. Next, we will get the location that not trusted through further analysis. Eventually meet people's expectations by improving.

Keywords: software trustworthiness, model checking, finite state machine model, trustworthy attribute

1 Introduction

Trust is essential to most human transactions [1]. Numerous research papers have addressed trust and software trustworthiness from many kinds of different perspectives in recent years [2]. However, at present, the existing researches mainly focus on two aspects, which are software reliability metrics and safety assessment [3]. M. Ohba divided the software reliability model into two categories: static model and dynamic model according to the modelling object [4]. Among them, the dynamic model becomes popular and has the most researchers. It models with some data or information related to the running time. This type of dynamic model utilizes software-testing process to obtain the failure time or software failure frequency over a period of time to estimate the number of failures of the entire software and time of failure occurrence or some other data involved with software failure. This typical model is Markov Process Model [5], Non-homogeneous Poisson Process (NHPP) [6] and Bayesian Model [7]. Certainly, there are also other extended models.

In addition, software interactivity cannot be neglected any longer because of that, too many safety issues are introduced through interaction. Nonetheless, people still do not keep a watchful eye on the measurement of the software interactivity. Therefore researching on software interactive security measure is a necessary complement for software reliability measure research and also a new development. This paper will model the software interaction as the state model and utilize the model-checking algorithm, which is a formal verification by exhaustively searching the finite state automata. And then convert the verification of properties to the corresponding temporal logic, using the model checking tool to traverse system model automatically, at last, it will check whether the system meets the corresponding properties or not. Compared with ordinary artificial validation method, model checking is of speed and high accuracy and is very useful for realizing the automation. The most important is that this model-checking algorithm not merely can reflect the behaviours of the software from the angle of function layer, but also further measure the software credibility from a performance perspective. Hence, this paper selects the model-checking algorithm of the IMC model to give two extended measurement to determine the software trustworthiness.

The scope of this paper is organized as follows: Section 2 introduces related work. In Section 3 presents the model checking in detail, especially the two extended measurement methods based on IMC and will utilize the two extended methods to decide the software trustworthiness. At the end of this chapter we verify the feasibility and effectiveness of these methods by experiments. Conclusions and some directions for future research are given in Section 4. Section 5 is the acknowledgements.

---

* *Corresponding author* e-mail: phc0409@126.com

2 Related work

Many reliability models and measurement methods have been proposed to estimate the software trustworthiness up to now. However, due to the software is more complex, so now there is no authority measurement in the world. At present, the most popular methods of software reliability analysis and evaluation include based on the development model, based on the informal method, based on the software behaviour, based on the formal method and based on the model checking measurement.

Software reliability analysis and measurement based on the development model usually makes full use of various development model to guarantee the reliability of the software. [8] proposed a trusted software design and development process based on model-driven architecture (MDA) which combines the executable formal specification language with UML description method to realize the executable formal specification description in the whole software development life cycle and guarantee the credibility of the software. This method can effectively detect the software behaviour to identify whether is trustable or not. Nonetheless, there is no detailed implement process and clear instructions. [9] described the software architecture applying AC2-ADL (Architectural Description of Aspect-Oriented Systems) and proposed a kind of trusted software architecture design method supporting run-time monitoring. This way can effectively achieves the trusted software system development process, but still need to further improve and research on software credible guarantee mechanism. [10] extended the trusted chain suggested by TCG (Trusted Computing Group) based on trusted computing platform. Through the description of irregular track, it inserted the corresponding check sensor into the key code that needed to be checked to implement dynamic reliable detection at the runtime. It is based on trusted computing and has the characteristics of high formalization, but the applicable scope is small.

Based on the formalization of software reliability mainly uses artificial method to analyse software and obtain the corresponding measure matrix to evaluate software reliability. [11] puts forward to extract different attribute benchmark index to evaluate the trusted degree based on the layered mechanism. This method is applicable to large modular software system. Nevertheless, the process of classifying the software reliable properties and obtaining the corresponding indicators is not fully automated. [12] proposed a trusted software process assessment method based on the

evidence. This way picks the objective data as the evaluation data. However, the corresponding metrics and algorithm still need further improvement.

In information security, the research on the behaviour of the software has always been used in intrusion detection. The theory of based on the software behaviour has been increasingly used in the dynamic measurement of trusted computing. [13] introduced a dynamic credible measurement based on software behaviour. At the same time, it puts forward an authentication mechanism based on expanded behaviour trace and behaviour measurement information. [14] used the behaviour track and checkpoints scenario to describe the dynamic characteristics, its aim is to detect the attacks. The software will stop running as long as finding any behaviour that deviation from the original expected track. Based on dynamic credible measurement, the measurements are divided into trusted or untrusted, but the credibility of software cannot be simply represented by trusted or untrusted.

Compared with the general software reliability analysis methods, the formal method based on strict mathematical foundation can carry on the formal descriptions or verification accurately and is suitable for reliability analysis and evaluation of the software.

Model checking is a kind of effective formal verification method. With the increasing development of model test technology, more and more researchers will apply the model checking technique to property verification of the code. [15, 16] both adopted the model checking method to validate the software trustworthiness. However, the current study is centred around the UML diagram of the early stage of the software development phase, for this reason, it does not go deep into the interaction level and also cannot verify the complex software trustworthiness in the operation phase. In this paper, we utilize the model-checking algorithm of IMC to extend the measurement of software trustworthiness.

3 Model checking

In this section, we will introduce the model-checking algorithm in detail. Next, the experiment, analysis and measurement will be given.

The structure of the model is roughly divided into three parts: modelling phase, running phase, analysis phase. The overall structure framework is shown in Figure 1 below.
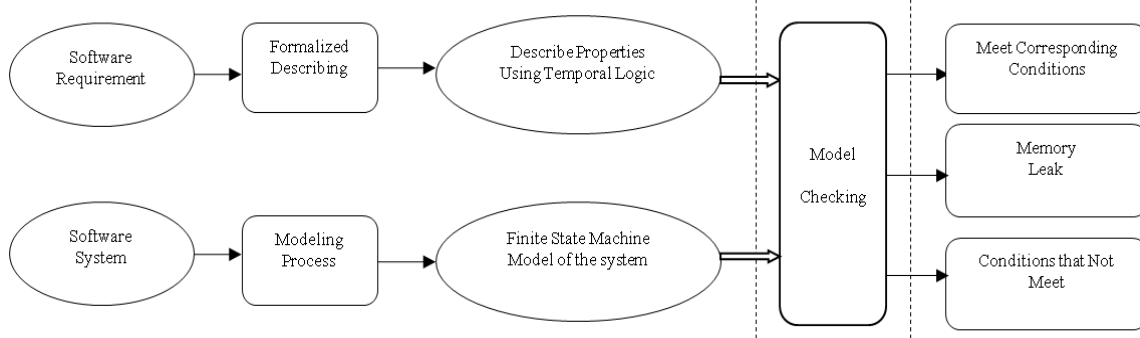
FIGURE 1 Overall structure framework

3.1 MODEL CHECKING ALGORITHM

The key algorithm is the two numerical iterative algorithm $F(s,t)$ and $G(s,t)$.

Theorem one: for $\varphi = \Phi_1\,_A U^{<t}\Phi_2$.

1. If $s \models \Phi_2$, $\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\Phi_2) = 1$.

2. If $(s \models \Phi_1 \wedge \neg\Phi_2) \wedge (s \in PS)$,

$\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\Phi_2) =$

$\sum_{s' \in S}\int_0^t R(s,s')e^{-E(s)x}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-x}\Phi_2)dx$.

3. If $(s \models \Phi_1 \wedge \neg\Phi_2) \wedge (s \in NS)$,

$\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\Phi_2) =$

$\sum_{R(s,s')\geq 0}\int_0^{\delta_A(s)} R(s,s')e^{-E(s)x}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-x}\Phi_2)dx +$

$\sum_{I(s,s')\in A}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-\delta_A(s)}\Phi_2)$.

4. $\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\Phi_2) = 1$.

Theorem two: for $\varphi = \Phi_1\,_A U^{<t}\,_B\Phi_2$.

5. If $(s \models \Phi_1) \wedge (\exists s'((s' \models \Phi_2) \wedge (\delta_B(s) \leq \delta_{A\backslash B}(s) < t)))$,

$\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\,_B\Phi_2) = 1$.

6. If $(s \models \Phi_1) \wedge (s \in PS)$,

$\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\,_B\Phi_2) =$

$\sum_{s' \in S}\int_0^t R(s,s')e^{-E(s)x}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-x}\,_B\Phi_2)dx$.

7. If $(s \models \Phi_1) \wedge (s \in NS) \wedge (\delta_{A\backslash B}(s) < \delta_B(s) < t)$

$\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\,_B\Phi_2) =$

$\sum_{R(s,s')\geq 0}\int_0^{\delta_{A\backslash B}(s)} R(s,s')e^{-E(s)x}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-x}\,_B\Phi_2)dx +$

$\sum_{I(s,s')\in A\backslash B}\mathrm{Prob}(s',\Phi_1\,_A U^{<t-\delta_{A\backslash B}(s)}\,_B\Phi_2)$.

8. $\mathrm{Prob}(s,\Phi_1\,_A U^{<t}\,_B\Phi_2) = 1$.

3.2 THE TWO EXTENDED MEASUREMENT METHODS

The model-checking algorithm can only assess the performance of the system. However, most users hope to know whether the software system is trusted or not and the measurement value. In addition, state transition as well as the parameters in the model is also important. In this part, we will give the two extended methods to solve the above problems.

3.2.1 The first extension measure

The first extended measure is to establish multiple corresponding temporal logic relations for each software trustworthy attribute that affecting software trustworthiness, that is to say that using multiple temporal logic formulas to describe a software trustworthy attribute, which is aim to measure the software trustworthiness on the multi-level and fine-grained. Then the paper will determine the measurement ultimately.

Here, we have to explain the trustworthy attribute, which generally refers to the functionality, the maintainability, the reliability, the survivability and the controllability of the software. Utilizing these attributes to describe the software trustworthiness. However, each attribute is expressed by multiple temporal logic formulas. The concrete practices are as follows.

The main algorithm of model checking just checks whether each state meet the path formula that is given. If meet the formula, it will return yes, whereas return no. In our first extension measure, we still adopt it. In addition, each temporal logic formula that corresponding to each trustworthy attribute will be taken into account. Then establish the corresponding relations between the important states and the results of these states whether meet each temporal logic formula, as shown below:

$$\{s_0, s_1, s_2, ..., s_n\} \xrightarrow[corresponds]{} \begin{bmatrix} \Phi_{00} & \Phi_{01} & \cdots & \Phi_{0m} \\ \Phi_{10} & \Phi_{11} & \cdots & \Phi_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ \Phi_{n0} & \Phi_{n1} & \cdots & \Phi_{nm} \end{bmatrix},$$

where $\Phi_{ij}(i \in n; j \in m)$ represents that the state s_i whether meet the jth temporal logic formula or not. If met, then: $\Phi_{ij}(i \in n; j \in m) = 1$ or $\Phi_{ij}(i \in n; j \in m) = 0$. It is easy to find that $\forall i, \forall j, \Phi_{ij}(i \in n; j \in m) = 1$ is the best condition.

Now assume that: $S = [s_1, s_2, ..., s_n]$, $s_i(i \in n)$ indicates the weight of state s_i . $L = [l_1, l_2, ..., l_m]$, $l_j(j \in n)$ indicates the weight of the jth temporal logic formula.

Then the last measurement of the whole software can be simply represented by $M_k = S_{1n} \cdot \Phi_{nm} \cdot L_{1m}^T$, M_k indicates the kth trustworthy attribute that affecting the software trustworthiness. After all are calculated, we can continue to choose the weighted average method to calculate the system reliability value. During this process, the most important step is to determine the corresponding temporal logic formulas for each trustworthy attribute, because only then can we really reflect the software system in detail. The experiment will be given in Section 3.3.

3.2.2 The second extension measure

The second extension measure is to locate the untrusted states, then find out the detail path and detail parameters of the path. Next, we will get the location that not trusted through further analysis. Eventually meet people's expectations by improving.

In the model of IMC, state transition is used to describe the path parameters. And the crucial factor of state transition is the occurrence time of acts and the state of residence time. Suppose we get the times, then we can clearly depict the system. Hence, the paper obtains the runtime parameters as follows:

$$\sigma = s_i < con_k, \delta(s_i), \delta_{act}(s_i) > s_j ... < ... > s_n,$$

where con_k denotes the jump from state s_i to state s_j belongs to the kth condition, $\delta(s_i)$ denotes the residence time of the state s_i , $\delta_{act}(s_i)$ denotes the occurrence time of the *act* from the state s_i .

The experiment will also be shown at the next section.

3.3 EXPERIMENT AND MEASUREMENT

In this part, we will give the related experimental data and analysis for the two extension measure.

3.3.1 The experiment of the first extension measure

The experiment example is the example of 6.5.1 in [17]. Figure 2 is the IMC model diagram, as follows:
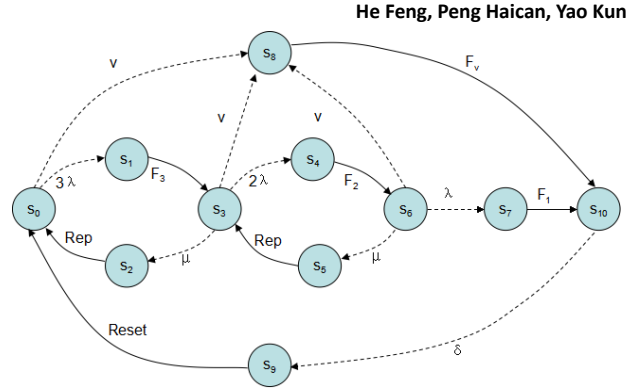
FIGURE 2 IMC model of a fault-tolerant system

Act is the set of acts: $Act = \{F_1, F_2, F_3, F_v, \text{Rep}, \text{Reset}\}$, F_i shows the ith processor is not work, F_v shows the vector cannot work normally Rep represents the fix act, the act of Reset can reset the system.

$$\lambda = 0.01, \quad \mu = 0.02, \quad \nu = 0.001, \quad \delta = 0.2,$$
$$\delta_{F_2} = 0.2, \quad \delta_{F_3} = 0.1, \quad \delta_{F_v} = 0.4, \quad \delta_{\text{Rep}} = 0.1,$$
$\delta_{\text{Reset}} = 0.1$. As shown in Figure 2, there is no doubt that the most important question is the fault tolerance in a fault-tolerant system of IMC model. However, the fault tolerance belongs to reliability. Then we select the reliability to describe the trustworthiness of the system. Temporarily we ignore other properties in this example. Next, we can use several temporal logic equation to describe the fault-tolerant system for fine-grained, as follows:

$$\Phi_1 = P_{<0.02}(true_{\{F_1, F_v\}} U^{<12}_{\{F_1, F_v\}} true),$$

$$\Phi_2 = P_{>0.2}(ture_{\varphi} U^{<12}_{\{F_1, F_2, F_3, \text{Re}\,p\}} true),$$

$$\Phi_3 = P_{>0.5}(true_{Act} U^{<12} \Phi_2).$$

The fault tolerance is depicted using the three temporal logic formulas. The first step is to build a relationship according the result as follows:

$$\{s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\} \underset{corrensponds}{\rightarrow}$$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

The second step is to determine the weight set of each state and the temporal logic formula respectively.

$$S = \{s_0, s_1, ..., s_{10}\} =$$
$$\{0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.06, 0.1, 0.12, 0.02\},$$

$$L = [l_1, l_2, l_3] = [0.3, 0.3, 0.4].$$

The reliability metrics is finally determined using the following formula:

$$M_k = S_{1n} \cdot \Phi_{nm} \cdot L_{1m}^T (n = 11, m = 3) = 0.6380 .$$

He Feng, Peng Haican, Yao Kun

For this particular system, the measurement of the fault tolerant is equal to the measurement of the reliability. So the trustworthiness measurement value of the fault-tolerant system is 0.6380. However, in general, the software trustworthiness is depicted by many trustworthy attributes. At this point, we should apply different method to synthesize according to the different system and situation.

We can conclude that temporal logic formulas data in table 1 and the corresponding results in table 2 according to the first extended method and algorithm procedures.

TABLE 1 Temporal logic formulas data

Prob(s_i,φ)	Φ_1	Φ_2	Φ_3
S_0	0.017	0.321	1.000
S_1	0.022	1.000	1.000
S_2	0.017	1.000	1.000
S_3	0.021	0.213	1.000
S_4	0.110	1.000	1.000
S_5	0.021	1.000	1.000
S_6	0.121	0.187	0.141
S_7	1.000	0.000	0.475
S_8	1.000	0.000	0.462
S_9	0.018	0.000	1.000
S_{10}	0.011	0.000	0.501

TABLE 2 Corresponding results

Φ_n	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
Φ_1	1	0	1	0	1	0	1	0	0	1	1
Φ_2	1	1	1	1	1	1	0	0	0	0	0
Φ_3	1	1	1	1	1	1	0	0	0	1	1

3.3.2. The experiment of the second extension measure

The way of the second extension measure is to obtain the detail path and path parameters for the states, which cannot meet the temporal logic formulas in the first extension measure. Then locate the positions that make the reliability low and give the reasons by analysing the path and the path parameters.

Here we still choose the example above, track the path and extract the operation path parameters to the fault-tolerant. We find the state s_6 does not meet the temporal logic formula 2 according the Table 2. Now we will track the path and extract the operation path parameters to the state s_6 as follows:

S6<0.3023,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.3810,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.2832,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.3551,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.2654,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.3331,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.2446,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.3086,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.2314,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.2785,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.2119,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.2608,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.1928,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.2376,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.1769,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.2181,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.1589,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.1961,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.1417,Maxdouble>-S5<Maxdouble,0.1>-

S3<0.1379,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.1288,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.1226,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.1477,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.1070,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.0984,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.1246,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.0869,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.0840,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.0766,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.0767,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.0687,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.0835,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.0585,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.0706,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.0459,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.0546,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.0340,Maxdouble>-S5<Maxdouble,0.1>-
S3<0.0394,Maxdouble>-S2<Maxdouble,0.1>-
S0<0.0250,Maxdouble>-S1<Maxdouble,0.1>-
S3<0.0290,Maxdouble>-S4<Maxdouble,0.2>-
S6<0.0150,Maxdouble>-S8<Maxdouble,0.4>-
S10<0.0014,Maxdouble>-S9<Maxdouble,0.1>

Among them, the *Maxdouble* represents the maximum double time. Here, if there is no Markov transfer except action transfer, we assume that the residence time of the state is *Maxdouble*. Similarly, if there is no action transfer except Markov transfer, we suppose the occurrence time of the action is *Maxdouble*.

From the above path, we can find that when the system start run from *s6* to *s9*, then the residual execution time is 0.0625 unit of time. However, the act starting from *s9* is only a Reset action operation, and the execution time of the Reset action is 0.1 unit of time. Hence, the vector fails due to the remaining time 0.0625 is less than 0.1. As a result, the state s_6 cannot meet the temporal logic formulas. Next, it affects the software trustworthiness and makes the measurement low.

4 Conclusions

This paper mainly proposed two expended measurement methods based on IMC model. The first expended method can give a final credibility value according to the result of temporal logic formulas. More than that, the result is intuitive and easy to understand to users. The second expended method can track the path and extract the operation path parameters for the important and untrusted states in accordance with the specific results of the first expended method. Of course, the intention is to analysis the cause of the result.

Software interaction is one of the most important key factors to the software reliability research. In the current open network environment, the introduction of interaction often leads to unpredictable risks, while this article on the basis of software interaction has proposed two extended methods, but there are a lot of limitations in this kind of methods based on IMC model. Moreover, the factors we considering are still not enough. So next, we want to introduce more data information on the basis of the dynamic interaction model, for example, the data or information that has nothing to do with the running time to

dynamically reflect the trustworthiness more accurately, comprehensively and truly.

The purpose to study the trusted software is to build the trusted software system that can meet the users, which requires the creditability validation before putting it into use. However, the current measurement theory, models and methods are mostly stay in theory, there are also some scholars that tried to apply various measurement methods in different kinds of industrial production, service system ,such as in [19-22], and obtained a series of research achievements, this article only carried on the analysis and verification of the experiment on a small fault

tolerance system. Then the focus of next step is to try to apply the extended methods to more areas of different systems. This will make the theoretical model can be used in real life successfully and embody the research significance.

Acknowledgments

References

[1] Cerf 2010 V G Trust and the Internet *IEEE Internet Computing* **14**(5) 95-6
[2] Yuyu Yuan and Qiang Han 2011 A Software Behavior Trustworthiness Measurement Method based on Data Mining, International Journal of Computational Intelligence System **4**(5) 817-25
[3] Tang Y, Liu Z 2010 Progress in Software trustworthiness metrics models *Computer Engineering and Applications* **46**(27) 12-6
[4] Ohba M 1984 Software reliability analysis models *IBM Journal of Research and Development* **28**(4) 428-43
[5] Littlewood B A 1975 Reliability model for systems with Markov structure *Applied Statistics* **24** 172-7
[6] Huang C Y, Lyu M R, Kuo S Y 2003 *IEEE Transactions on Software Engineering* **29**(3) 261-9
[7] Littlewood B, Verrall J L 1973 A Bayesian reliability growth model for computer software *Applied Statistics* **22**(3) 332-46
[8] Tang Y, Du Y, Liu W 2009 Design of Trusted Software Based on MDA and Executable Formalization *Computer Engineering* **35**(19) 138-40
[9] Wen J, Wang H, Ying S, Ni Y, Wang T 2010 Toward a Software Architectural Design Approach for Trusted Software Based on Monitoring Chinese Journal of Computers **33**(12) 2321-34 *(in Chinese)*
[10] Tian J, Li Z, Liu Y 2011 A Design Approach of Trustworthy Software and Its Trustworthiness Evaluation *Journal of Computer Research and Devolopment* **48**(8) 1447-54 *(in Chinese)*
[11] Mukherjee A, Siewiorek D P 1997 *IEEE Transactions on Software Engineering* **23**(6) 366-78
[12] Du J, Yang Y, Wang Q, Li M 2011 Evidence-Based Trustworthy Software Process Assessment Method *Journal of Frontiers of Computer Science and Technology* **6** 501-12 *(in Chinese)*
[13] Zhuang L, Cai M, Li C 2010 Software Behavior-Based Trusted Dynamic Measurement *Wuhan University (Nat Sci Ed)* **56**(2) 133-7 *(in Chinese)*
[14] Cheng L, Zhang Y 2009 A Verification Method of Security Model Based on UML and Model Checking *Chinese Journal of Computers* **32**(4) 1035-1039 *(in Chinese)*
[15] He F, Zhang H, Yan F, Yang Y, Wang H, Meng X 2010 Test of Trusted Software Stack Based on Model Checking *Wuhan University (Nat Sci Ed)* **56**(2) 129-32 *(in Chinese)*
[16] Wu J, Wu Y, Tan G 2007 Interactive Markov Chain: The Design, Verification and Evaluation of Concurrent System *Science Press: Beijing (in Chinese)*
[17] Zhuang L, Cai M, Shen C 2011 Trusted Dynamic Measurement Based on Interactive Markov Chains *Journal of Computer Research and Development* **48**(8) 1464-72 *(in Chinese)*
[18] Mohammed M H,·Lim C P,·Quteishat A 2014 A novel trust measurement method based on certified belief in strength for a multi-agent classifier system *Neural Computing and Applications* **24**(2) 421-9
[19] Huynh T D, Jennings N R, Shadbolt N R 2006 Developing an Integrated Trust and Reputation Model for Open Multi-Agent Systems *Autonomous Agents and Multi-Agent Systems* **13**(2) 119-54
[20] Saint Germain B, Valckenaers P, Van Belle J, Verstraete P, Van Brussel H 2012 Incorporating trust in networked production systems *Journal of Intelligent Manufacturing* **23**(6) 2635-46
[21] Zhan G, Shi W, Deng J 2009 Sensor Trust: A Resilient Trust Model for Wireless Sensing Systems *ACM Sensys* Ann Arbor USA 1-40

Authors

Feng He, born in November, 1964, Ningxia, China

Current position, grades: full professor of Computer Science at Computer Department, Jiaxing College, China.
University studies: M.Sc. in Mathematics (1977), PhD in Computer Sciences (2008) at Donghua University, China.
Scientific interest: Database and knowledge engineering, service-oriented computing
Experience: Lead and participated in National and Provincial Scientific research projects.

Haican Peng, born in April, 1987, Henan, China

Current position, grades: master's degree student.
University studies: Computer Science.
Scientific interest: information system analysis and integration, data mining.

Kun Yao, born in February, 1989, Shanxi, China

Current position, grades: Hanzhong Branch Company of Shanxi Province of China Telecom.
University studies: master's degree in Computer Technology (2013).
Scientific interest: computer networks and information security.

A nonlinear system modelling approach to industrial cane sugar crystallization

Guancheng Lu, Yanmei Meng*, Jian Chen, Zhihong Tang, Xiaochun Wang, Xian Yu

College of Mechanical Engineering, Guangxi University, No. 100, Daxue Road, Nanning, China

Received 1 July 2014, www.cmnt.lv

Abstract

Cane sugar crystallization is a non-linear process where multiple control parameters are involved, which makes it rather difficult to reveal its internal mechanism by mechanism modelling. Derived from variants of standard support vector machine method, an online control system modelling method based on multi-input and multi-output proximal least square support vector machine is proposed to be applied in sugar crystallization process. This method takes multiple process control parameters as the input and output of machine learning algorithm, through which the inherent law between key and auxiliary parameters in the sugar crystallization process is established. The ultimate goal is to control the sugar crystallization process automatically. The experimental results show that the accuracy rate of the model output is 95%.

Keywords: multi-input and multi-output proximal least square SVM, sugar crystallization control system, machine learning, nonlinear modelling

1 Introduction

Nowadays techniques like soft sensor based on computer and sensing technology, process identification and nonlinear process control are widely applied in modern industrial field. Among them, process identification and soft sensor based on computer sensing technology is a method that establishes the mathematic relationship between leading variables and auxiliary variables, where the leading variables are those hard to be measured or even can't be measured and the auxiliary variables are those relevant and easy to be measured. Then the evaluation of leading variables is implemented by auxiliary variables. An organic combination of computer and industrial process knowledge has been achieved by soft sensor technology, which makes variables measurement easier and more accessible by replacing hardware with software and taking auxiliary variables to evaluate the leading ones.

Cane sugar crystallization is a nonlinear and complex process as well as hard to be modelled. The key parameters in that control process field like super saturation and Brix are not stable enough for long term measurement. The accuracy of measuring sensors would be compromised once their surfaces are covered with scaling sugar in crystallization process, which stands in the way of automatic controlling in Chinese sugar industry field. However, machine learning methods could identify the inherent law between the input and output in control process through learning and modelling the relevant data, without knowing exactly what kind of mechanism the process is. It means that the machine learning methods can approach to the actual process closely.

Support Vector Machine (SVM) is a new machine learning technique based on minimization principle, with features of solid theoretical foundation and basis [1-3]. SVM which features with good generalization ability can still obtain forecast value under some extreme conditions, like nonlinearity, small sample and high-dimension data [4-6]. SVM is generally applied in industrial fields of iron and steel production, biopharming and food processing for parameter forecasting, with trend of penetrating into other industrial fields.

2 Measurement and control system model for sugar crystallization process

Crystallization is the most significant phase of sugar boiling, and its mechanism is complex and hard to be modelled. The key control parameters like massecuite Brix and supersaturation can be obtained through measuring and analysing the following external parameters: temperature, vacuum degree, steam pressure etc. Since those external parameters determine the massecuite Brix and supersaturation, there should exist some certain law between them. Even though the inherent law can be established by analysing the mechanism of sugar crystallization process, it is still difficult to figure it out from the perspective of the mechanism modelling since multiple parameters are involved and the process is nonlinear.

Since the internal mechanism of sugar crystallization process is hard to be modelled based on traditional theory, this paper applies machine learning to cave the implicit relation among the key control parameters in sugar

---

* *Corresponding author* e-mail: gxu_mengyun@163.com

crystallization process. After that, the measurement and control system model for sugar crystallization process has been built, shown as Figure 1.
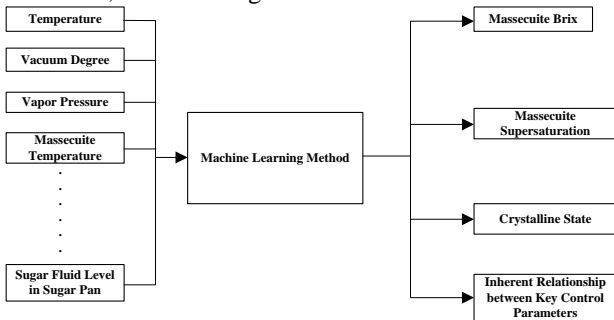
FIGURE 1 Control modelling scheme of cane sugar crystallization process

3 Base modelling building

In this paper, machine-learning method was adopted to extract the inherent law between the multiple key parameters and the auxiliary parameters in the nonlinear cane sugar crystallization process, which aimed to controlling the process automatically. The multiple control parameters in that process were seen as the input and output of the machine learning method.

The input vector $x(x \ni R^d)$ in d dimensions was defined as the auxiliary parameter in the process; whereas the output vector $Y(Y \ni R^n)$ in n dimensions was defined as the key parameter. Measurement and control system

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}_j} = \mathbf{w}_j - \sum_{i=1}^{m} a_{i,j} \Phi_j(\mathbf{x}_i) = 0 \rightarrow \mathbf{w}_j = \sum_{i=1}^{m} a_{i,j} \Phi_j(\mathbf{x}_i) \\ \dfrac{\partial L}{\partial b_j} = b_j - a_{i,j} = 0 \rightarrow b_j = a_{i,j} \\ \dfrac{\partial L}{\partial e_{i,j}} = ce_{i,j} = a_j = 0 \rightarrow e_{i,j} = \dfrac{a_{i,j}}{c} \\ \dfrac{\partial L}{\partial a_{i,j}} = 0 - \left\{ \mathbf{w}_j^T \Phi_j(\mathbf{x}_i) + b_j + e_{i,j} - y_{i,j} \right\} = 0 \rightarrow \mathbf{w}_j^T \Phi_j(\mathbf{x}_i) + b_j + e_{i,j} - y_{i,j} = 0 \\ i = 1,2,...,m \\ j = 1,2,...\,n \end{cases} \tag{3}$$

Linear equations system of eliminating the transition variables in Equation (3):

$$\begin{bmatrix} \Phi(\mathbf{x}_1)\Phi(\mathbf{x}_1)+1+\dfrac{1}{c} & \cdots & \Phi(\mathbf{x}_1)\Phi(\mathbf{x}_m)+1 \\ \vdots & \ddots & \vdots \\ \Phi(\mathbf{x}_m)\Phi(\mathbf{x}_1)+1 & \cdots & \Phi(\mathbf{x}_m)\Phi(\mathbf{x}_m)+1+\dfrac{1}{c} \end{bmatrix} \times$$

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} = Y, \tag{4}$$

model for cane sugar crystallization process was built based on multi-input multi-output proximal least square support vector machine (MIMO PLS-SVM) [7, 8]. The object function for optimizing the model satisfies the following equation:

$$\min_{w_j, e_{i,j}} J^{(n)}\left(\mathbf{w}_j, e_{i,j}\right) = \frac{1}{2}\sum_{j=1}^{n}\left(\mathbf{w}_j^T\mathbf{w}_j + b_j^2\right) + \frac{c}{2}\sum_{j=1}^{n}\sum_{i=1}^{m} e_{i,j}^2, c<0,$$

$$s.t \begin{cases} y_{i,1} = \mathbf{w}_1^T\Phi_1(\mathbf{x}_i) + b_1 + e_{i,1}, \; i=1,2,...,m \\ y_{i,2} = \mathbf{w}_2^T\Phi_2(\mathbf{x}_i) + b_2 + e_{i,2}, \; i=1,2,...,m \\ ... \\ y_{i,n} = \mathbf{w}_n^T\Phi_n(\mathbf{x}_i) + b_n + e_{i,n}, \; i=1,2,...,m \end{cases} \tag{1}$$

where b_j is the j-dimensional offset vector; w_j is the j-dimensional weight vector; x is the high dimensional transformation function; $e_{i,j}$ is the quantitative error in i^{th} row and j^{th} column; c is the penalty factor of irrelevant samples. The following Lagrange function is deduced from the object function:

$$L^{(n)}\left(\mathbf{w}, b_j, e_{i,j}, a_{i,j}\right) =$$
$$J^{(n)} - \sum_{j=1}^{n}\sum_{i=1}^{m} a_{i,j}\left\{\mathbf{w}_j^T\Phi_j(\mathbf{x}_i) + b_j + e_{i,j} - y_{i,j}\right\}, \tag{2}$$

The conditions for optimality are the following equations:

According to the principle of SVM, kernel function is introduced as follows:

$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1,\mathbf{x}_1) & \cdots & K(\mathbf{x}_1,\mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_m,\mathbf{x}_1) & \cdots & K(\mathbf{x}_m,\mathbf{x}_m) \end{bmatrix}, \tag{5}$$

$$a = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}, \tag{6}$$

$$E = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}, \qquad (7)$$

$$I = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}. \qquad (8)$$

Consequently, the linear equations system can be written as matrix form:

$$\left(\mathbf{K} + \mathbf{E} + c\mathbf{I} \right)\mathbf{a} = \mathbf{Y}. \qquad (9)$$

Left Matrix in Equations (9) is named as model identification matrix as follows:

$$\mathbf{H} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) + 1 + \dfrac{1}{c} & \cdots & K(\mathbf{x}_1, \mathbf{x}_m) + 1 \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_m, \mathbf{x}_1) + 1 & \cdots & K(\mathbf{x}_m, \mathbf{x}_m) + 1 + \dfrac{1}{c} \end{bmatrix}. \qquad (10)$$

Introduced by the positive definiteness of the kernel function, we can get the following relationship:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j) \right) = \left(\Phi(\mathbf{x}_j)\Phi(\mathbf{x}_i) \right) = K(\mathbf{x}_j, \mathbf{x}_i),$$

which suggest matrix H is symmetric. Assuming z is a nonzero vector, do the operation as follows:

$$\mathbf{z}^T \mathbf{H} \mathbf{z} = \mathbf{z}^T \mathbf{K} \mathbf{z} + \mathbf{z}^T \mathbf{E} \mathbf{z} + c\mathbf{z}^T \mathbf{I} \mathbf{z} =$$

$$\left\| \sum_{i=1}^{l} \lambda_i \Phi(\mathbf{x}_i) \right\|^2 + \left(\sum_{i=1}^{l} z_i \right)^2 + \frac{1}{c} \sum_{i=1}^{l} z_i^2 > 0. \qquad (11)$$

$$\mathbf{H}_{t+1} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) + 1 + \dfrac{1}{c} & \cdots & K(\mathbf{x}_1, \mathbf{x}_t) + 1 & K(\mathbf{x}_1, \mathbf{x}_{t+1}) + 1 \\ \vdots & \ddots & \vdots & \vdots \\ K(\mathbf{x}_t, \mathbf{x}_1) + 1 & \cdots & K(\mathbf{x}_t, \mathbf{x}_t) + 1 + \dfrac{1}{c} & K(\mathbf{x}_t, \mathbf{x}_{t+1}) + 1 \\ K(\mathbf{x}_{t+1}, \mathbf{x}_1) + 1 & \cdots & K(\mathbf{x}_{t+1}, \mathbf{x}_t) + 1 & K(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) + 1 + \dfrac{1}{c} \end{bmatrix}. \qquad (13)$$

Assuming:

$$\mathbf{h}_{new} = \begin{bmatrix} K(\mathbf{x}_{t+1}, \mathbf{x}_1) + 1 \\ \vdots \\ K(\mathbf{x}_{t+1}, \mathbf{x}_t) + 1 \end{bmatrix}, \qquad (14)$$

$$h_{new} = K(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) + 1 + \frac{1}{c}. \qquad (15)$$

Since matrix H is symmetric and positive, Equation (9) has a unique solution due to the presence of its inverse matrix.

Known from the object function, the measurement and control model built on MIMO PLS-SVM can feed the quantified correlativity among multiple variables in the cane sugar crystallization process back to the output variables efficiently and synchronously. Mutual adjustment between those outputs variables are made on condition of the identical input variables, which leads to coordination of the whole control objects. In a certain sense, the whole measurement and control process is optimized.

4 Online measurement and control model building of cane sugar crystallization process

According to Equations (9) and (10), the solution of Equation (9) is determined by the inverse matrix of the symmetric and positive H. However, solving the inverse matrix directly is of high computation cost and time-consuming, which is incompatible with the online measurement and control model. Obviously, matrix inversion approach available for online measurement and control condition is highly demanded.

At time t, matrix H is defined as follows:

$$\mathbf{H}_t = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) + 1 + \dfrac{1}{c} & \cdots & K(\mathbf{x}_1, \mathbf{x}_t) + 1 \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_t, \mathbf{x}_1) + 1 & \cdots & K(\mathbf{x}_t, \mathbf{x}_t) + 1 + \dfrac{1}{c} \end{bmatrix}. \qquad (12)$$

At time $t + 1$, the new data described as input vector x_{t+1} is inserted to the online model, then H can be written as follows:

According to the inversion of block matrix when new data is inserted to the online model, the inversion of H_{t+1} is obtained using the inverse of H_t at time t, which simplifies the computation process massively [9]. At time $t + 1$, the solution of linear equations system of the online model is defined as a_{t+1}; the newly added data is regarded as input vector y_{t+1}; Y_{t+1} denotes the output matrix; Consequently a_{t+1} can be rewritten by Equation (17), where the needed computing time is decreased largely. Equations system of online measurement and control model is assumed as $H_{t+1}a_{t+1}=Y_{t+1}$, and the solution deduction is as follows:

Mathematical and Computer Modelling

$$
\mathbf{H}_{t+1}^{-1} = \begin{bmatrix} \mathbf{H}_t & \mathbf{h}_{new} \\ \mathbf{h}_{new}^T & h_{new} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{H}_t^{-1} + \mathbf{H}_t^{-1}\mathbf{h}_{new}^T\mathbf{H}_t^{-1}\left(h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T\right)^{-1} & \mathbf{H}_t^{-1}\mathbf{h}_{new}\left(h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T\right)^{-1} \\ \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\left(h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T\right)^{-1} & \left(h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T\right)^{-1} \end{bmatrix} =
$$
$$
\begin{bmatrix} \mathbf{H}_t & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{H}_t^{-1}\mathbf{h}_{new} \\ -1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_{new}^T\mathbf{H}_t^{-1} & -1 \end{bmatrix} \frac{1}{h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T}, \tag{16}
$$

$$
\mathbf{a} = \mathbf{H}_{t+1}^{-1}\mathbf{Y}_{t+1} = \left(\begin{bmatrix} \mathbf{H}_t & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{H}_t^{-1}\mathbf{h}_{new} \\ -1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_{new}^T\mathbf{H}_t^{-1} & -1 \end{bmatrix} \frac{1}{h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T} \right) \mathbf{Y}_{t+1} =
$$
$$
\begin{bmatrix} \mathbf{H}_t^{-1}\mathbf{Y}_t \\ 0 \end{bmatrix} + \frac{1}{h_{new} - \mathbf{h}_{new}^T\mathbf{H}_t^{-1}\mathbf{h}_{new}^T} \left(\begin{bmatrix} \mathbf{h}_{new}^T\mathbf{H}_t^{-1} & -1 \end{bmatrix} \mathbf{Y}_t - \mathbf{y}_{t+1} \right) \begin{bmatrix} \mathbf{H}_t^{-1}\mathbf{h}_{new} \\ -1 \end{bmatrix}, \tag{17}
$$

The data of the online model would increase rapidly along with time flowing and data updating, which compromises the operation speed. Aimed to solve this problem, the data amount needed processing must be limited without compromising the identification precision of the model. When the data amount reaches to a certain point, space for processing the newly data would be squeezed out by deleting the old data. In the meantime, the inverse matrix and solution of the equations system of the model would be updated. Sliding window techniques are used to clear the old data out the online model. Preprocessing is made as follows:

$$
\mathbf{h}_{old} = \begin{bmatrix} K\left(\mathbf{x}_2, \mathbf{x}_1\right)+1 \\ \vdots \\ K\left(\mathbf{x}_{t+1}, \mathbf{x}_1\right)+1 \end{bmatrix}, \tag{18}
$$

$$
h_{old} = K\left(\mathbf{x}_1, \mathbf{x}_1\right) + 1 + \frac{1}{c}. \tag{19}
$$

Then we have:

$$
\mathbf{H}_{t+1}^{-1} = \begin{bmatrix} h_{old} & \mathbf{h}_{old} \\ \mathbf{h}_{old}^T & \bar{\mathbf{H}}_{t+1} \end{bmatrix}^{-1} = \begin{bmatrix} \left(h_{old} - \mathbf{h}_{old}\bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{old}^T\right)^{-1} & -\mathbf{h}_{old}^T\bar{\mathbf{H}}_{t+1}^{-1}\left(h_{old} - \mathbf{h}_{old}^T\bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{p;d}^T\right)^{-1} \\ -\bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{old}^T\left(h_{old} - \mathbf{h}_{old}^T\bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{old}^T\right)^{-1} & \bar{\mathbf{H}}_{t+1}^{-1} + \bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{old}^T\mathbf{h}_{old}\bar{\mathbf{H}}_{t+1}^{-1}\left(h_{old} - \mathbf{h}_{old}\bar{\mathbf{H}}_{t+1}^{-1}\mathbf{h}_{old}^T\right)^{-1} \end{bmatrix} \tag{20}
$$

$$
\mathbf{H}_{t+1}^{-1} = \begin{bmatrix} h_{inverse\_old} & \mathbf{h}_{inverse\_old} \\ \mathbf{h}_{inverse\_old}^T & \mathbf{H}_{inverse\_new} \end{bmatrix}. \tag{21}
$$

Compare Equation (20) with (21), the inverse matrix of cane sugar measurement and control model is as follows, when the old data is clear at time $t+1$:

$$
\bar{\mathbf{H}}_{t+1}^{-1} = \mathbf{H}_{inverse\_new} - \frac{\mathbf{h}_{inverse\_old}^T\mathbf{h}_{inverse\_old}}{h_{inverse\_old}}, \tag{22}
$$

Equations (23) and (24) determine the solution of the model, whose deductive process is as follows:

$$
\mathbf{a}_{t+1} = \mathbf{H}_{t+1}^{-1}\mathbf{Y}_{t+1},
$$

$$
\begin{bmatrix} \mathbf{a}_{old} \\ \hat{\mathbf{a}}_{t+1} \end{bmatrix} = \begin{bmatrix} h_{inverse\_old} & \mathbf{h}_{inverse\_old} \\ \mathbf{h}_{inverse\_old}^t & \mathbf{H}_{inverse\_new} \end{bmatrix} \times \begin{bmatrix} \mathbf{Y}_{old} \\ \bar{\mathbf{Y}}_{t+1} \end{bmatrix},
$$

$$
\mathbf{a}_{old} = h_{inverse\_old}\mathbf{Y}_{old} + \mathbf{h}_{inverse\_old}\bar{\mathbf{Y}}_{t+1}, \tag{23}
$$

$$
\hat{\mathbf{a}}_{t+1} = \mathbf{h}_{inverse\_old}^T\mathbf{Y}_{old} + \mathbf{H}_{inverse\_old}\bar{\mathbf{Y}}_{t+1}. \tag{24}
$$

When the old data is replaced by newly data at time $t+1$, the solution of model is defined as $\bar{\mathbf{a}}_{t+1}$ while the output vector is $\bar{\mathbf{Y}}_{t+1}$. Then the solution of model can be determined by Equation (25), through which the computing process is simplified and large-scale matrix operation is avoided.

$$
\bar{\mathbf{a}}_{t+1} = \bar{\mathbf{H}}_{t+1}^{-1}\bar{\mathbf{Y}}_{t+1} \Rightarrow
$$

$$
\bar{\mathbf{a}}_{t+1} = \left(\mathbf{H}_{inverse\_new} - \frac{\mathbf{h}_{inverse\_old}^T\mathbf{h}_{inverse\_old}}{h_{inverse\_old}} \right) \times \bar{\mathbf{Y}}_{t+1} \Rightarrow
$$

$$
\bar{\mathbf{a}}_{t+1} = \mathbf{H}_{inverse\_new}\bar{\mathbf{Y}}_{t+1} + \mathbf{h}_{inverse\_old}^T\mathbf{Y}_{old} - \mathbf{h}_{inverse\_old}^T\mathbf{Y}_{old} - \frac{\mathbf{h}_{inverse\_old}^T\mathbf{H}_{inverse\_old}\mathbf{h}_{inverse\_old}^T\bar{\mathbf{Y}}_{t+1}}{h_{inverse\_old}} \Rightarrow
$$

$$
\bar{\mathbf{a}}_{t+1} = \mathbf{H}_{inverse\_new}\bar{\mathbf{Y}}_{t+1} + \mathbf{h}_{inverse\_old}^T\mathbf{Y}_{old} - \mathbf{h}_{inverse\_old}^T\frac{h_{inverse\_old}\mathbf{Y}_{old} + \mathbf{h}_{inverse\_old}\bar{\mathbf{Y}}_{t+1}}{h_{inverse\_old}},
$$

$$\overline{\mathbf{a}}_{t+1} = \hat{\mathbf{a}}_{t+1} - \frac{\mathbf{h}_{inverse\_old}^{T} \mathbf{a}_{old}}{\mathbf{h}_{inverse\_old}} \ . \tag{25}$$

The online identification process is as follows:
1) Acquire a little sample data from the control or detection process. Optimize the kernel function parameter and penalty factor of the model, aiming to obtain the optimal parameters.
2) The parameters of model are offline trained in order to build the offline model.
3) The input data of model is received and identified online.
4) The actual control or measurement process determines if new data is inserted to the data set of model [10]. If new data is inserted in, add the data to the data set, otherwise the new data is abandoned and jump to step (3).
5) Set the size of data set, once its size is larger than the limited one, delete the original data and jump to step (3).

If a table is too long to fit onto one page, the table number and headings should be repeated on the next page before the table is continued.

Alternatively, the table can be spread over two consecutive pages (first on even-numbered, then on odd-numbered page).

For a wide table you can use 1-column section (Table 1), for a small standard table 2-column section is used (Table 2).

5 Simulation and experimental analysis

This paper takes the data provided by a local cane sugar production factory as experiment object. Data selected from multiple cane sugar crystallization phases is defined as experiment sample set to make sure that the experiment can cover the whole range of sugar crystallization process and data for machine learning is sufficient. The sample set has 212 samples, and parts of them are shown as Table 1. It is a common sense that the massecuite Brix and supersaturation are two vital factors in the cane sugar crystallization process. Massecuite temperature, steam pressure, steam temperature, vacuum degree and sorts of these are defined as input vectors, whereas massecuite Brix and supersaturation are defined as output vectors. The measurement and control model of the cane sugar crystallization process can synchronously identify the massecuite Brix and supersaturation online. It can also compare the results with the measured Brix.

TABLE 1 Experimental data

Sample number	Brix (°Bx)	Supersaturation (%)	Massecuite temperature (°C)	Vacuum degree (kPa)	Steam temperature (°C)	Steam pressure (MPa)
1	81.75	1.18	58.91	82.56	108.9	0.068
2	79.18	1.10	64.36	84.56	109.2	0.065
3	81.69	1.21	57.85	81.36	107.5	0.036
4	79.91	1.09	59.81	82.32	108.2	0.03
5	83.97	1.25	58.06	80.4	108.2	0.052
6	82.39	1.19	59.67	83.4	106.5	0.037
7	81.62	1.14	56.78	83.04	106.9	0.038
…	…	…	…	…	…	…
212	78.37	1.06	67.65	83.16	109.8	0.063

Gaussian radial basis function is selected as the kernel function of the model talked about above. The experiment program is compiled In VC9.0 environment. Parameters and penalty factors of the kernel function are optimized using particle swarm optimization (PSO) algorithm and leave-one-out cross validation [11]. Assume the optimization range of parameters was from 0.01 to 1000, and the penalty factors' range was from 1 to 10000. 162 samples are selected randomly from the sample set as

training set and the rest of sample set is regarded as testing set (50 samples). Iterative algebra of PSO is 150 and population size of it is 100. Emulation results are shown in Table 2. Aimed to verify the experiment results, the difference between the measured data and output data from the model has been studied, which is shown in Figures 2 and 3 shows the accuracy rate of the results; experimental errors were demonstrated as Figure (4).

TABLE 2 Results

Sample number	Brix (°Bx)	Supersaturation (%)	Massecuite temperature (°C)	Vacuum degree (kPa)	Error rate of Brix (%)	Error rate of supersaturation (%)
1	78.17	1.06	78.69	1.09	0.67	2.83
2	78.12	1.05	78.09	1.07	0.04	1.90
3	85.1	1.26	84.78	1.28	0.37	1.59
4	78.07	1.06	79.40	1.09	1.70	2.83
5	79.07	1.07	80.03	1.09	1.22	1.87
6	83.26	1.21	81.51	1.2	2.10	0.83
7	82.81	1.2	81.90	1.21	1.10	0.83
…	…	…	…	…	…	…
50	78.91	1.1	79.06	1.12	0.19	1.82
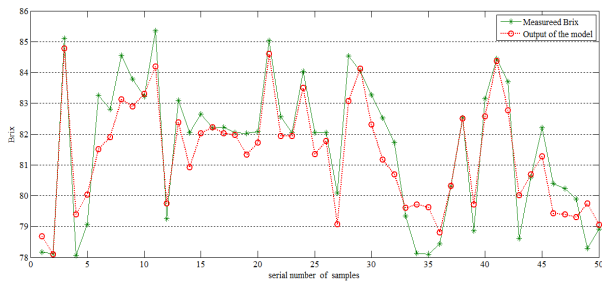
FIGURE 2 Comparative results of Brix
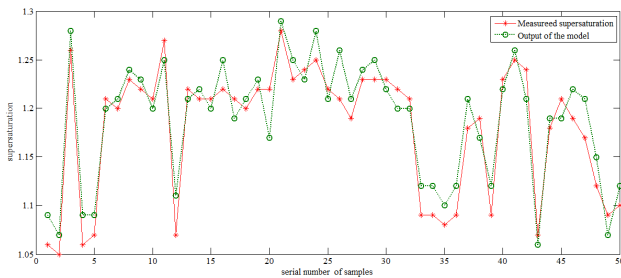
FIGURE 3 Comparative results of supersturation
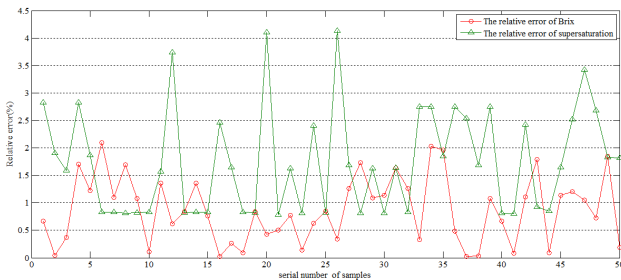
FIGURE 4 The relative error of Brix and supersaturation

6 Conclusion

Modelling with MIMO PLS-SVM can still perceive the predictive value of high accuracy when numbered samples are trained, by which challenges of nonlinearity, numbered samples and high dimension are easily dissolved. The inherent relationship between the multiple key variables and auxiliary variables is established with application of machine learning. The auxiliary variables are defined as the input of machine learning algorithm, while the key variables are defined as the output. Experimental results demonstrated that the accuracy rate of output was 95%. In an overall sense, the model discussed above is of great utility value in process identification, soft sensor and nonlinear control field.

Acknowledgement

References

[1] Suykens J A K, Lukas L, Vandewalle J 2000 Sparse approximation using least square vector machine *Proceedings of the ISCAS 2000 Geneva The 2000 IEEE International Symposium on Circuits and Systems* **2** 757-60
[2] Suykens J A K, Lukas L, Vandewalle J 2002 Weighted least squares support vector machines: Robustness and sparse approximation *Neurocomput* **48**(1-4) 85-105
[3] Mountrakis G, Im J, Ogole C 2011 Support vector machines in remote sensing: A review *ISPRS Journal of Photogrammetry and Remote Sensing* **66**(3) 247-59
[4] Chang C C, Lin C 2011 LIBSVM: A Library for Support Vector Machine *ACM Transactions on Intelligent Systems and Technology* **2**(3) Article No. 27
[5] Kim H S, Sohn S Y 2010 Support vector machines for default prediction of SMEs based on technology credit *European Journal of Operational Research* **201** 838-46

[6] Kamandar M, Ghassemian H 2013 Linear Feature Extraction for Hyperspectral Images Based on Information Theoretic Learning *IEEE Geoscience and Remote Sensing Letters* **10**(4) 702-6
[7] Charrada A, Samet A 2012 SIMO-OFDM Channel Estimation based on Nonliear Complex LS-SVM *International Journal of Computer Applications* **43**(3) 1-7
[8] Sun Y, Todorovis S, Goodison S 2010 Local Learning Based Feature Selection for High Dimensional Data Analysis *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9) 1610-26
[9] Jiao L C, Bo L F, Wang L 2007 Fast Sparse Approximation for Least Squares Support Vector Machine *IEEE Transactions on Neural Networks* **18**(3) 685-697
[10] Olivieri M, Clinton J 2012 An almost fair comparisonbetween Earthworm and SeisComp3 *Seismol Res Lett* **83**(4) 720-7
[11] Bo L, Wang L, Jiao L C 2006 Feature scaling for kernel fisher discriminant analysis using leave-one-out cross validation *Neural Comput* **18**(4) 961-78

Authors

Guancheng Lu, born in 1984, China

Current position, grades: teacher and engineer at Guangxi University.
University studies: Master degree in mechanical and electronic engineering, Guangxi University, China, 2010.
Scientific interest: measurement and control system modelling, automatic control implementation for mechanical engineering.
Publications: 3 Patents, 6 papers.

Yanmei Meng, born in 1963, China

Current position, grades: PhD, professor, vice dean of the mechanism and engineering college in Guangxi University. China.
University studies: PhD degree in Mechanical Manufacturing and Automation in Hefei University of Technology.
Scientific interest: Complex mechatronic product design and optimization, intelligent detection and control technology.
Publications: 11 Patents, 60 Papers.

Jian Chen, born in 1989, China

Current position, grades: postgraduate student in Guangxi University.
University studies: Bachelor degree in Mechatronic Engineering in 2012.
Scientific interest: model predictive control, PLC implementation.
Publications: 1 paper.

Zhihong Tang, born in 1964, China

Current position, grades: senior engineer in college of Mechanism engineering of Guangxi University.
University studies: Bachelor degree in Mechatronic Engineering.

Xiaochun Wang, born in 1954, China.

Current position, grades: associate professor at the college of Mechanism engineering of Guangxi University.
University studies: Bachelor degree in Mechanical manufacturing in 1982.
Scientific interest: advanced manufacturing technology, industrial engineering.
Publications: 20 Papers.
Experience: 20 years in Guangxi University.

Xian Yu, born in 1989, China

Current position, grades: postgraduate student in Guangxi University.
University studies: bachelor degree in Mechatronic Engineering in 20.12.
Scientific interest: artificial intelligent control, model optimization and data mining, first-principle and hybrid modeling of sugar crystallization.
Publications: 1 Patent, 2 Papers.

Finite element modelling and modal analysis of structure with bolted joints

Liping He[1*], Limiao Qian[1], Zijing Wang[1], Kai Long[2]

[1]*Beijing Institute of Radio Measurement, Beijing, China*

[2]*State Key Laboratory for Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, China*

Abstract

In order to investigate a modal analysis of structure with bolted joints, five kinds of finite element models, a solid bolt model, a no-bolt model, a modified bolt model, a beam bolt model and a rigid bolt model, are introduced. Among these models, the solid bolt model, modified bolt model and rigid bolt model provide good accurate responses compared with the modal test results. The solid bolt model, which can also be used for stress analysis, is recommended in view of applicability. The rigid bolt model is recommended in view of usefulness and effectiveness. For bolted flanged connection, the magnitude of bolt pretension has little influence to structural modes. To improve the calculation accuracy of higher modes, a better idea is to modify the finite model by modal test.

Keywords: bolted joints, flange, pretension, modal, finite element analysis

1 Introduction

A bolted joint is a kind of mechanical joint being widely used in engineering structures. Two primary characteristics in the bolted joint are a pretension and a mating part contact, which have great influence to both static and dynamic characteristics of structure. The study on modelling and simulation of bolted joint has important engineering application and theoretical research value.

To analysing the connection stiffness and contact stress of bolted joint, a large number of studies have been carried out in theory, by experiment and FEM (Finite Element Method) [1-5]. It is verified that the pressure cone exists in members around the bolt hole. Yujuan Sun et al. studied the axial-load and stress distributions on each thread of the threaded connection based on axisymmetric and 3-D finite element models [6, 7]. Simulation results indicate that the effect of helix has little influence to axial-load and stress distributions of threaded connection. These studies provide a theoretical support to simplify the bolt model. Jeong Kim et al. introduced four kinds of finite element models for bolted joint [8], which can be used for stress analysis of members. The modified bolt model can also be used for modal analysis. DayuPu et al. proposed another two models for modal analysis [9]. However, all of these models have limited application scope because they cannot be applied to stress analysis and modal analysis simultaneously.

In order to investigate a modal analysis of bolted flanged connection, five kinds of finite element models, a solid bolt model, a no-bolt model, a modified bolt model, a beam bolt model and a rigid bolt model, are introduced based on finite element software ANSYS. Among these

models, the solid bolt model, modified bolt model and rigid bolt model provide good accurate responses compared with the modal test results. The solid bolt model, which can also be used for stress analysis, is recommended in view of applicability. The rigid bolt model is recommended in view of usefulness and effectiveness.

2 Basic principle of prestressed modal analysis

Theoretical analysis and engineering practice have shown that the damping has little influence to structural natural frequency and mode shape. So damping can be ignored when solving the structure natural frequency and mode shape. Natural frequency and mode shape are inherent attributes of structural system, and have nothing to do with the external load. The structural non-damping free vibration equation can be expressed as

$$[M]\{\ddot{U}\}+[K]\{U\}=\{0\}, \tag{1}$$

where $[M]$ is the mass matrix, $[K]$ is the stiffness matrix, $\{\ddot{U}\}$ is the acceleration vector, $\{U\}$ is the displacement vector.

For a linear system, the free vibration is of harmonic form

$$\{U\}=\{\phi\}_i \cos\omega_i t, \tag{2}$$

where $\{\phi\}_i$ is the eigenvector, that is mode shape, corresponding to the 1st, 2nd,…, ith natural frequency, ω_i is the ith circular frequency, t is the time.

Corresponding author e-mail: bithlp@sina.com

Substituting Equation (2) into Equation (1) gives

$$\left([K] - \omega_i^2 [M]\right)\{\phi\}_i = \{0\} . \tag{3}$$

The circular frequency ω_i and mode shape $\{\phi\}_i$ can be obtained by solving Equation (3).

In order to consider the influence of bolt pretension to structural modes, ANSYS preforms nonlinear static analysis before modal analysis to obtain actual contact status, contact stiffness and prestress, from which the system stiffness matrix $[K]$ can be updated. The subsequent modal analysis uses the updated stiffness matrix, so as to ensure the accuracy of modal analysis [10].

3 Finite element model of bolted flanged connection

Figure 1 shows a typical bolted flanged connection structure. The upper flange and lower flange are bolted with 8 groups of M10 bolts distributed circumferential uniform. Figure 2 shows the details of bolted joint. In order to investigate a modal analysis of bolted flanged connection, five kinds of finite element models are introduced. The first one, that is the solid bolt model, takes into account the actual bolt pretention and contacts, but the rest four bolt models are simplified models, without considering the bolt pretention and contacts.

3.1 SOLID BOLT MODEL

The solid bolt model as shown in Figure 1 is the most realistic finite element model among them. Hexahedral elements are adopted to establish the detailed models of bolt, nut, washer and flange. Without considering the threads, the bolt and nut are connected together by coupling nodes. The contacts between nut and upper washer, bolt head and lower washer, upper washer and upper flange, lower washer and lower flange, upper flange and lower flange are simulated by surface-to-surface contact model. Pretention elements are inserted into cross section of stud, so as to apply precise pretention load. Figure 2 shows the details of solid bolt model, which include five contact pairs and a group of pretention elements.

3.2 NO-BOLT MODEL

The no-bolt model as shown in Figure 3 is the simplest finite element model among them. Without considering bolt, nut and washer, this model is composed of upper flange and lower flange. The actual contact between upper flange and lower flange is ignored, but modelled by coupling nodes.
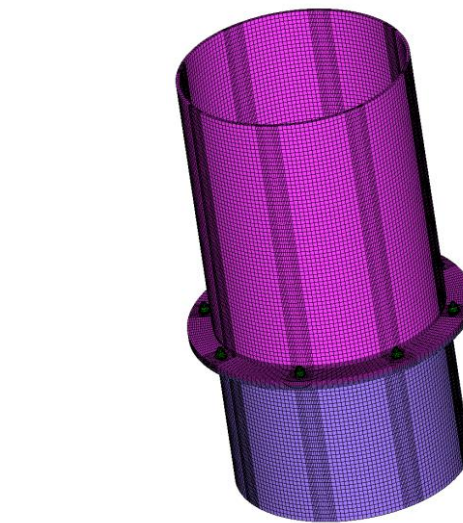
FIGURE 1 Typical bolted flanged connection structure

Contact elements

Pretention elements
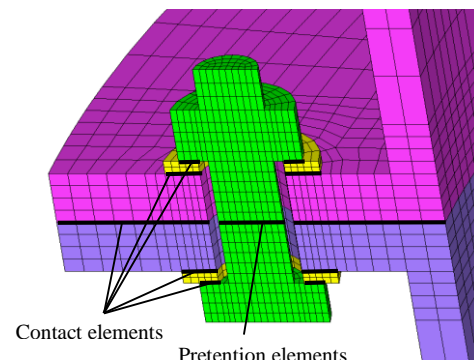
FIGURE 2 Solid bolt model

Merged
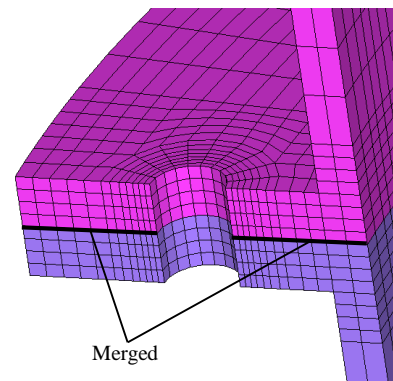
FIGURE 3 No-bolt model

3.3 MODIFIED BOLT MODEL

Earlier studies had found that larger stress, which holds all parts together, exists around the area of bolted joint due to the clamping force of bolt pretention. The solid bolt model is modified according to Osgood's suggestion [11]. All the contact pairs and pretention elements are deleted. The bolt, nut, washer and flange are connected together by coupling nodes as shown in Figure 4. It is noted that the upper flange

and lower flange are not connected together except in the annular region shown in Figure 4.
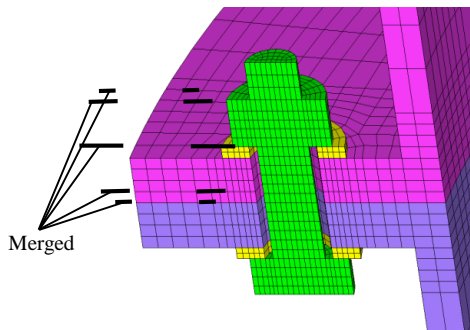
FIGURE 4 Modified bolt model

3.4 BEAM BOLT MODEL

Figure 5 shows the beam bolt model. Flanges, belong to plates, are modelled by shell elements. The bolt head, nut and washer are ignored. The stud is approximately modelled by a beam element, and the nodes at both ends of the stud are connected to the flanges respectively by rigid elements in a spider-web-like style. The contact area of washer and flange is a rigid region because of the rigid elements. This approach is effective since the number of finite elements is significantly reduced as well as the ignorance of bolt pretention and contact compared with the solid bolt model.
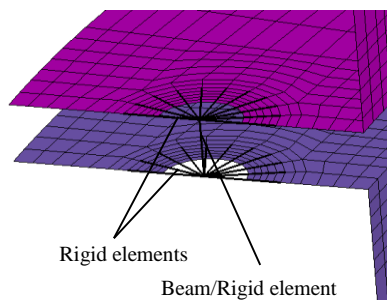
FIGURE 5 Beam/Rigid bolt model

3.5 RIGID BOLT MODEL

To replace the beam element by rigid element for the stud, the beam bolt model changes to rigid bolt model (shown in Figure 5), which has higher computing efficiency.

4 Modal analysis and test

The modal analysis of bolted flanged connection belongs to prestressed modal analysis. The bolt pretention and contact behaviour are included in solid bolt model, but the nonlinear static analysis should be executed before modal analysis. Without considering bolt pretention and contact

behaviour, the rest four bolt models can execute modal analysis directly.

The modal test is carried out with single input and multi outputs modal testing method based on LMS modal test system. The test structure is hung from a soft string so as to simulate an unconstrained state as shown in Figure 6. Four acceleration sensors are used for measuring the response of vibration successively to reduce the influence of added mass to structural modes. 144 points are chosen in the test structure to obtain the accurate mode shapes as shown in Figure 7. The influence of bolt pretention to structural modes can be studied by two modal tests. The pretention load, 2KN, is applied to each bolt for the first test, and the load is 10KN for the second test. The bolt pretention can be applied by torque wrench. The relationship of tightening torque and bolt pretention can be expressed as

$$T = 0.2F_0 d , \qquad (4)$$

where T is the tightening torque, F_0 is the bolt pretention, d is the nominal diameter of bolt.
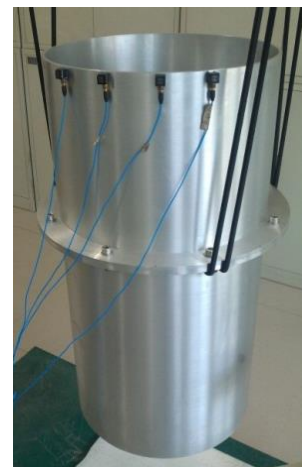
FIGURE 6 Free modal test
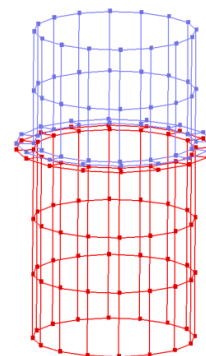
FIGURE 7 Location of measuring points

Table 1 lists the computing efficiency for each bolt model. Table 2 lists the natural frequencies obtained from modal test and finite element analysis. Figure 8-11 show the mode shapes obtained from modal test and solid bolt model (The bolt pretention is 10KN).

He Liping, Qian Limiao, Wang Zijing, Long Kai

TABLE 1 Computing efficiency for each bolt model

Models	No. of nodes	No. of elements	Efficiency
Solid bolt model	156568	128408	Low
No-bolt model	121320	88544	Medium
Modified bolt model	151480	111232	Medium
Beam/Rigid bolt model	35860	35240	High

TABLE 2 Natural frequencies of bolted flanged structure (Hz)

Mode no.	1	3	5	7
Experiment when F_0=2KN	107.55	265.10	290.49	408.87
Experiment when F_0=10KN	107.65	264.70	290.69	408.14
Solid bolt model when F_0=2KN	105.20	268.45	283.57	357.82
Error	-2.28%	1.42%	-2.45%	-12.33%
Solid bolt model when F_0=10KN	105.20	268.46	283.57	357.83
Error	-2.28%	1.42%	-2.45%	-12.33%
No-bolt model	108.83	289.82*	287.37*	469.90
Error	1.10%	9.49%	-1.14%	15.13%
Modified bolt model	105.74	269.88	284.22	360.14
Error	-1.77%	1.96%	-2.23%	-11.76%
Beam bolt model	105.22	244.29	283.17	331.43
Error	-2.26%	-7.71%	-2.59%	-18.80%
Rigid bolt model	105.50	264.83	283.81	355.15
Error	-2.00%	0.05%	-2.37%	-12.98%

Notes: the 3rd and 5th mode shapes obtained from no-bolt model are corresponding to the 5th and 3rd mode shapes obtained from test respectively. The two data with * in Table 2 have been switched as to correspond to mode shapes obtained from test. The error is the deviation of simulation and test data when F_0=10KN.

Simulation and experiment both can obtain repeated roots due to the symmetry of structure. The 2nd, 4th, 6th, 8th mode shapes are repeated roots of the 1st, 3rd, 5th, 7th mode shapes respectively. Repeated roots have the same frequency and mode shape, but have different phases. Repeated roots are not listed in Table 2 and analysed in next sections.

Experiment and solid bolt model both indicate that the bolt pretention has little influence to structural natural frequency and mode shape. The contact status between upper flange and lower flange changes from the whole surface contact into local contact due to bolt pretention. Figure 12 and Figure 13 show the contact status between upper flange and lower flange when the bolt pretention is different. The red annular areas around the bolt holes in Figure 12 and Figure 13 are contact areas which are almost identical. It indicates that the upper flange and lower flange are held together around the bolt holes. The magnitude of bolt pretension only changes the local connection stiffness, but has little influence to structural frequency and mode shape.

Experiment and simulation data both show that the solid bolt model, modified bolt model and rigid bolt model provide good accurate responses. The deviations of the simulation and test data remain less than 2.5% for the 1st, 3rd, 5th natural frequencies. The 3rd natural frequency calculated by beam bolt model which provides lower connection stiffness is lower by 7.71% compared with experiment data. The 3rd natural frequency calculated by no-bolt model which overestimates connection stiffness between upper flange and lower flange is higher by 9.49% compared with experiment data.

For the 7th mode, mode shapes obtained from all kinds of bolt models are the same to experiment result, but the deviations of frequency are more than 10% for every model. In fact, the near contact region (yellow area in Figure 12) between upper flange and lower flange may contact each other while vibration, which affects the higher modes more significantly. For modal analysis, the solid bolt model can simulate the contact status and stiffness after applying bolt pretention. However, the contact status and stiffness will not change any more no matter how the structure vibrates. The rest four bolt models can't consider connection stiffness change without bolt pretention and contact pair modelled. To improve the calculation accuracy of higher modes, a better idea is to modify the finite model by modal test.
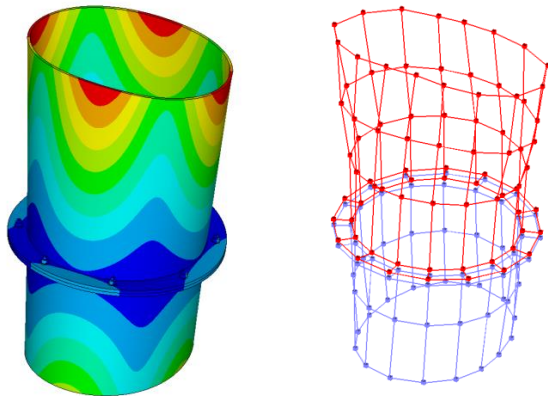
a) Simulation (105.20Hz)   b) Test (107.65Hz)

FIGURE 8 1st mode shape
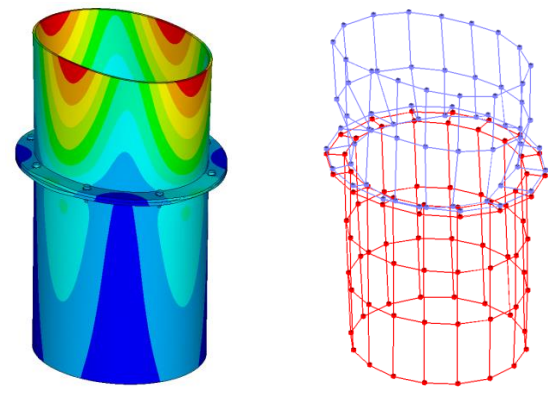
a) Simulation ((268.46Hz)   b) Test (264.70Hz)

FIGURE 9 3rd mode shape
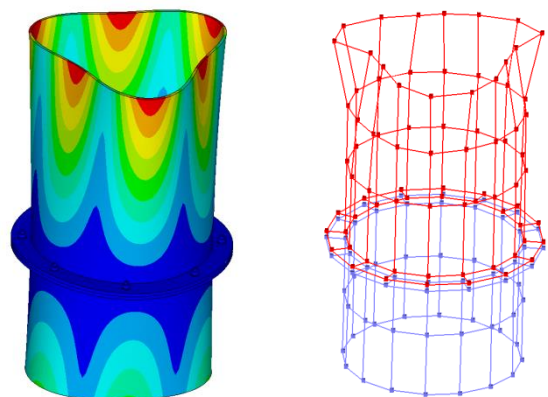
a) Simulation (283.57Hz)   b) Test (290.69Hz)

FIGURE 10 5th mode shape
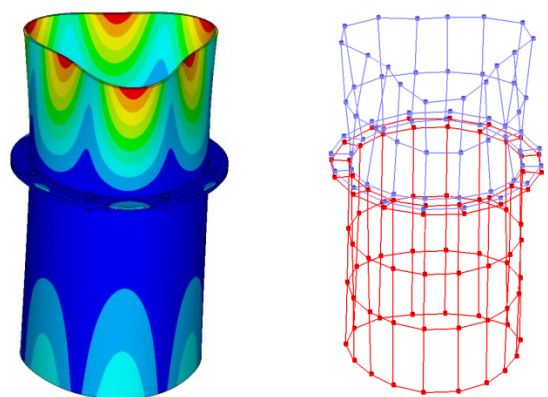
a) Simulation (357.83Hz)   b) Test (408.14Hz)

FIGURE 11 7th mode shape
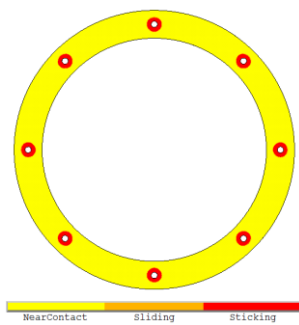
FIGURE 12 Contact status between flanges for F_0 =2KN
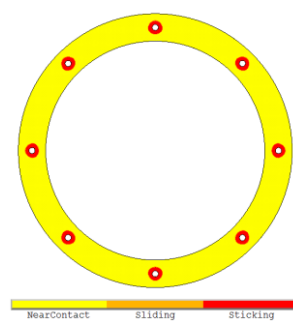
FIGURE 13 Contact status between flanges for F_0 =10KN

5 Conclusions

In order to investigate a modal analysis of structure with bolted joints, five kinds of finite element models, a solid bolt model, a no-bolt model, a modified bolt model, a beam bolt model and a rigid bolt model, are introduced. Among these models, the solid bolt model, modified bolt model and rigid bolt model provide good accurate responses compared with the modal test results. The solid bolt model, which can also be used for stress analysis, is recommended in view of applicability. The rigid bolt model is recommended in view of usefulness and effectiveness.

Experiment and simulation both indicate that the bolt pretention has little influence to structural natural frequency and mode shape. The contact status between upper flange and lower flange changes from the whole surface contact into local contact due to bolt pretention. The magnitude of bolt pretension only changes the local connection stiffness, but has little influence to structural frequency and mode shape.

The near contact region between upper flange and lower flange may contact each other while vibration, which affects the higher modes more significantly. All the bolt models can't simulate changing contact status and connection stiffness. To improve the calculation accuracy of higher modes, a better idea is to modify the finite model by modal test.

References

[1] Bradly T L, Lardner T J, Mikic B B 1971 *Transactions of the ASME* 542-5

[2] Gould M M, Mikic B B 1972 *Transactions of the ASME* 864-70

[3] Wileman J, Choudhury M, Green I 1991 *Transactions of the ASME* 432–7

[4] Lehnhoff T F, Ko K II, Mckay M L 1994 *Transactions of the ASME* 550-7

[5] Nassar S A, Abboud A 2009 *Journal of Mechanical Design* 121001-1-121001-11

[6] Sun Y J, Liao R D 2011 *Advanced Materials Research* 148-149 1741-4

[7] Sun Y J, Liao R D, Zhang W Z 2009 *Chinese Journal of Mechanical Engineering* **22**(6) 869-75

[8] Kim J, Yoon J C, Kang B S 2007 *Applied Mathematical Modelling* **31**(5) 895- 911

[9] Pu D Y, Liao R D, Zuo Z X, et al. 2009 *Journal of Beijing Institute of Technology* **18**(3) 273-7

[10] ANSYS Inc 2012 *ANSYS Release 14.5 Documentation Structural Analysis Guide*

[11] Osgood C C 1979 Saving Weight in Bolted Joints *Machine Design*

Authors

Liping He, born on August 23, 1982, Sichuan Province, China

Current position, grades: doctor of Philosophy, senior engineer in Beijing Institute of Radio Measurement.
University studies: PhD in Human-Computer & Environment Engineering from Beijing Institute of Technology (2005-2010).
Scientific interest: finite element method, dynamics modelling & simulation, structural optimization.
Publications: 3 patents, 10 papers, 1 book.

Limiao Qian, born on January 28, 1974, Hebei Province, China

Current position, grades: Master of Engineering, senior engineer in Beijing Institute of Radio Measurement.
University studies: M.E. in Vehicle Engineering from Beijing Jiaotong University (2000-2003).

Zijing Wang, born on March 12, 1986, Jilin Province, China

Current position, grades: master of engineering, assistant engineer in Beijing Institute of Radio Measurement.
University studies: B.E. in Engineering Mechanics & Automation from Beijing Jiaotong University (2005-2009), M.E. in Mechatronics from Beihang University (2009-2012).

Kai Long, born on August 28, 1978, Hubei Province, China

Current position, grades: doctor of philosophy, Lecture in North China Electric Power University.
University studies: M.E. and PhD in Power and Mechanical Engineering from Beijing Institute of Technology (2000-2007).
Scientific interest: finite element method, structural optimization.
Publications: 40 papers, 3 books.

An automatic approach to detecting and eliminating lazy classes based on abstract syntax trees

Wei Liu¹, Zhigang Hu¹, ²*, Hongtao Liu²

¹*School of Information Science and Engineering, Central South University, Changsha 410083, China*

²*School of Software, Central South University, Changsha 410075, China*

Abstract

To detect and eliminate lazy classes in source code, an automatic approach based on abstract syntax trees (ASTs) is proposed. Source code files transform to ASTs at first, then the relationships between classes are extracted from the ASTs. Three common relationships are considered, which are generalization, association and dependency. Some definitions are proposed to represent the classes set of different kinds of relationships. After carrying out several set operations on these sets, the candidate lazy classes set is obtained. By further manual examination, the true lazy classes are acquired. Finally, a specific lazy class will be removed automatically from the project. Four projects are tested to detect and eliminate the lazy classes. The experimental results show that the proposed detection algorithm has high precision rate. In addition, this approach has good efficiency, and its execution time has a linear relationship to the size of a system.

Keywords: code smells, lazy classes, abstract syntax trees, refactoring, classes set

1 Introduction

Code smells are symptoms or indicators in the source code that indicate potential problems. The well-known 22 code smells are described in [1] by Martin Fowler and Kent Beck. The identifications or detections of code smells are useful in the sense that they might constitute prescriptive guidance for performing certain types of refactoring. Some common code smells emerge frequently in the existing code, such as Duplicated Code, Long Method, Large Class, Lazy Class, Switch Statements and so on. Code smells affect the maintainability of software systems, and they are important indicators for code refactoring [2, 3].

Recently, code smells detection and automatic refactoring become hotspots in software engineering research. Lots of code smell detection approaches have been proposed. Radu Marinescu [4] presented a metric-based approach to detecting code smells with detection strategies and developed a PRODETECTION toolkit that supported code inspections based on detection strategies. Naouel Moha et al. [5, 6] proposed a DECOR (DEtection & CORrection) method that described all the steps necessary for the specification and detection of code and design smells. Moreover, they introduced an approach to automating the generation of detection algorithms from specifications written using a domain-specific language, and they specified 10 smells and generated automatically relevant detection algorithms using templates. Foutse Khomh et al. [7, 8] presented a Bayesian approach for the detection of code and design smells, their approach could handle the inherent uncertainty of the detection process. In

addition, they presented BDTEX (Bayesian Detection Expert) and GQM (Goal Question Metric) based approach to building Bayesian Belief Networks (BBNs) from the definitions of code smells. Hui Liu et al. [9] proposed a detection and resolution sequence for different kinds of bad smells to simplify their detection and resolution, they highlighted the necessity of managing bad smell resolution sequences with a motivating example and recommended a suitable sequence for commonly occurring bad smells. A. Ananda Rao et al. [10] proposed a quantitative method, which made use of the concept design change propagation probability matrix (DCPP matrix) to detect two important bad smells, which were shotgun surgery and divergent change. Although there are many code smells detection approaches, few of them can detect Lazy Class. So, a special approach to detecting and eliminating lazy classes is needed. Min Zhang et al. [11] performed a systematic literature review of 319 papers about code bad smells and analysed in detail 39 of the most relevant papers, they found that our knowledge of some code bad smells remains insufficient and some code bad smells receive little most research attention, such as the Lazy Class.

In addition, some tools have been developed for detecting code bad smells automatically and several research works have been done on them. Francesca Arcelli Fontana et al. [12, 13] gave reviews about the current panorama of the tools for automatic code smell detection, and they assessed many frequently-used tools, such as CheckStyle, in Fusion, PMD, and so on. Moreover, they outlined the main differences among these tools and the different results they obtained. Their research results show

---

Corresponding author e-mail: zghu@csu.edu.cn

that few tools could detect Lazy Class, let alone eliminate Lazy Class automatically.

2 Lazy class and abstract syntax tree

2.1 LAZY CLASS

Lazy Class is one of the bad smells in code, which indicates a useless class or a class with few responsibilities. Each class, which we have created should cost time and money to maintain and understand. Too many lazy classes will increase the complexity and scale of a software system. So, a class that is not doing enough to pay for itself should be eliminated. But if we face a huge project with millions of source code lines and thousands of classes, it is a so hard thing to find all lazy classes by manual handling. How to detect and eliminate lazy classes automatically is a meaningful topic in software engineering, especially in the field of code smells identifying and refactoring.

In this paper, we propose a novel and systematic approach to detecting and eliminating lazy classes automatically.

2.2 ABSTRACT SYNTAX TREE

In order to detect and eliminate lazy classes, we can analysis UML diagrams such as class diagrams. But a class diagram only describes high level relationships between classes, it loses some detail information, such as some dependency relationships. To get more relationship information between classes, we have to handle source code directly. However, source code analysis will raise the complexity and execution time. It has higher time and space complexity. In addition, during the stage of source code analysis, there is a lot of useless information affecting the execution efficiency.

To balance the complexity and efficiency for detecting the relationships between classes, we need a trade-off method. Abstract Syntax Tree (AST) is a proposed way to represent source code, which contains more information than class diagram. AST is used as an intermediate expression. In our approach, we use Java language as a sample, and the proposed approach can be used for other object-oriented languages. The Abstract Syntax Tree maps plain Java source code in a tree form, which is more convenient and reliable to analyse and modify programmatically than text-based source [14]. Every Java source file is entirely represented as tree of AST nodes that are all subclasses of the ASTNode.

In our approach, Eclipse is used as an IDE (Integrated Development Environment) to analysis Java source code and implement refactoring. It provides JDT (Java Development Tools) and Eclipse AST to handle Java source code. Eclipse JDT contains a group of APIs to access and operate source code, it contains two different ways to access Java source code: Java Model and AST. Eclipse AST is an important part of Eclipse JDT, which is defined in the package named org.eclipse.jdt.core.dom. In Eclipse AST, there are some classes to modify, create, read, and delete source code. In order to have good expandability and flexibility, Eclipse AST is designed based on the Factory Method pattern and the Visitor pattern [15].

3 Automatic detection and elimination algorithm

After transforming source code to abstract syntax trees by Eclipse AST, we can detect and extract all relationships between classes by handling the ASTs. If a class is a lazy class (redundancy class), it has no relationship to other classes. In general, there are three kinds of relationships between classes, including generalization, association and dependency. If we find that all of the other classes have no any relationship to a specific class, the class is maybe an islet. It means that the class is very likely a lazy class. So, the problem of detection of lazy classes is transformed to a problem of finding isolated classes.

In order to describe the process for searching isolated classes and detecting candidate lazy classes, a series of definitions are proposed as follows.

Definition 1: Project Classes Set (PCS). PCS is a set that stores all classes' names in a project.

Definition 2: Super Classes Set (SCS). SCS is a set that stores all super classes' names of a specific class. The super interfaces are also in SCS.

Definition 3: Associate Classes Set (ACS). ACS is a set that stores all associate classes' names of a specific class. Association classes' instances are attributes of a specific class.

Definition 4: Dependent Classes Set (DCS). DCS is a set that stores all dependent classes' names of a specific class. Generally, dependent relationships are represented by three main ways: a class's instance is one of the parameters of another class's method, a class's instance is the local variable in a method of another class, and a class invokes another class's static methods. If a class has one of the three aforementioned relationships to a specific class, it will be added to the specific class's DCS.

Definition 5: Relevant Classes Set (RCS). RCS of a class is a union set of the class's SCS, ACS and DCS. The formula to calculate RCS of class i as follows:

$$RCS(i) = SCS(i) \bigcup ACS(i) \bigcup DCS(i) . \tag{1}$$

In addition, RCS of a project is a union set of all classes' RCS in a software system. The equation to calculate RCS of a project as follows:

$$RCS(\text{Project}) = \bigcup_{i=1}^{n} RCS(i) . \tag{2}$$

Definition 6: Lazy Classes Candidate Set (LCCS). LCCS of a project is a set that stores all candidate lazy classes in a software system. Candidate lazy classes are in the *PCS* but not in the *RCS*. We can obtain the *LCCS*(Project) using the following equations:

Liu Wei, Hu Zhigang, Liu Hongtao

$$LCCS(\text{Project}) = PCS(\text{Project}) - RCS(\text{Project}) . \qquad (3)$$

For example, if the $RCS(\text{Project}) = \{A, B, C, E, F\}$ and the $PCS(\text{Project}) = \{A, B, C, D\}$, the $LCCS(\text{Project}) = PCS(\text{Project}) - RCS(\text{Project}) = \{A, B, C, D\} -g \{A, B, C, E, F\} = \{D\}$. Here, E and F are in RCS but not in PCS, which are called library classes, such as the classes in JDK

or other open source libraries. Library classes list in RCS, but they are not parts of the current system and do not list in the PCS. D is a candidate lazy class, it is a part of the system but maybe none of the others needs it Pseudo-code of automatic detection algorithm for candidate lazy classes is listed in Table 1.

TABLE 1 Pseudo-code of the automatic detection algorithm

Line	Pseudo-code
	Input: The name of the root directory, which contains source code files for detecting.
	Output: A set stored all candidate lazy classes' names.
1	declare a null Set named projectClassSet
2	declare a null Set named relevantClassSetofProject
3	
4	for each source code file in the directory
5	create an AST for the file
6	add the class name to projectClassSet
7	
8	declare a null Set named superClassSet
9	store all directly super classes of the current class to superClassSet
10	
11	declare a null Set named associateClassSet
12	declare a null Set named dependentClassSet
13	for each FieldDeclaration in the AST
14	store the field's type name (not primitive type) to associateClassSet
15	if there is a ClassInstanceCreation node
16	store the type name of the instance in ClassInstanceCreation node to dependentClassSet
17	end if
18	if there are TypeLiteral nodes
19	store the type names of classes in TypeLiteral nodes to dependentClassSet
20	end if
21	end for
22	
23	for each MethodDeclaration in the AST
24	store the parameters' type names (not primitive type) to dependentClassSet
25	store the exceptions' type names to dependentClassSet
26	store the type names of instances in all ClassInstanceCreation nodes to dependentClassSet
27	store the type names of classes in all static MethodInvocation nodes to dependentClassSet
28	store the type names of classes in all static fields access nodes (QualifiedName) to dependentClassSet
29	store the type names of exception in all CatchClause nodes to dependentClassSet
30	store the type names in all InstanceofExpression nodes to dependentClassSet
31	store the type names of classes in all TypeLiteral nodes to dependentClassSet
32	end for
33	
34	declare a null Set named relevantClassSetofClass
35	relevantClassSetofClass = superClassSet $\cup$ associateClassSet $\cup$ dependentClassSet
36	relevantClassSetofProject = relevantClassSetofProject $\cup$ relevantClassSetofClass
37	end for
38	
39	declare a null Set named lazyClassSet
40	lazyClassSet = projectClassSet - relevantClassSetofProject
41	return lazyClassSet

In Table 1, projectClassSet is used to store PCS (Line 1) and relevantClassSetofProject is used to store RCS of a project (Line 2). For each source code file in the project, an AST is created firstly, then the relevant class name is added to projectClassSet (Line 6). In Line 8-9, superClassSet is used to store SCS, and associateClassSet is declared to store ACS in Line 11 and dependentClassSet is declared to store DCS in Line 12. In Line 13-21, for each field of the class, if the type of field is not a primitive type, the type name is stored to associateClassSet which is used to store ACS. ClassInstanceCreation node and TypeLiteral node are also considered in the FieldDeclaration. If there

is a ClassInstanceCreation node or a TypeLiteral node, relevant class's name will be added to DCS. In Line 23-32, all dependent classes are extracted from each method, eight situations are considered to detect different kinds of dependent classes. At last, relevantClassSetofClass is used to store RCS of current class (Line 34), and relevantClassSetofClass is the union set of SCS, ACS and DCS (Line 35). The relevantClassSetofClass is added to relevantClassSetofProject (Line 36). In Line 39, a set named lazyClassSet is used to store LCCS, lazyClassSet is the difference between projectClassSet and

Liu Wei, Hu Zhigang, Liu Hongtao

relevantClassSetofProject, and the lazyClassSet is returned finally (Line 40-41).

Obviously, time complexity of the algorithm is $T(n) = n \times (m + k)$, here, n is the number of source code files in the project, m is the average number of fields in each class and k is the average number of methods in each class. Generally, m and k are not too large. If we define a suitable constant C, we can consider as: $1 \leq (m + k) \leq C$, and $T(n) = n \times (m + k) \leq C \times n = O(n)$. It means that the execution time has a linear relationship to the number of source code files, and the automatic detection algorithm has good efficiency.

After detecting all candidate lazy classes stored in LCCS, we have to examine the candidates meticulously by manual. Some candidate lazy classes are not true lazy classes, for example, the entry class of a system, or a class, which is located in a configuration file, or a class which is used in user interface files (e.g. Java Server Pages). If a real lazy class is confirmed, we should eliminate it automatically. Pseudo-code of the automatic elimination algorithm for a lazy class is listed in Table 2.

In Table 2, all source code files in the project are checked. Several files maybe have more than one class,

and each class transforms to a TypeDeclaration node respectively. In this approach, we do not save the package name of a class, so these classes which have a same class name should be considered. If another class's name equals to a specific class's name, corresponding TypeDeclaration node is stored into lazyClassNodeList. Finally, if we find that the size of lazyClassNodeList is greater than 1, it means that there are at least two classes with the same class name, we need to select the target class by manual. Otherwise, the TypeDeclaration node is deleted automatically. Time complexity of this automatic elimination algorithm is: $T(n) = n \times m$, here, n is the number of source code files in the project and m is the average number of class in each file. Most of the files have only one class, and a few files have more than one class. For a large project, we can assume that m trends to a constant. So, the time complexity is: $T(n) = O(n)$. Algorithm's execution time is proportional to the number of source code files, which can be used to represent the scale of system.

TABLE 2 Pseudo-code of the automatic elimination algorithm

Line	Pseudo-code
	Input: The name of the root directory, which contains source code files before refactoring and the class name of a true lazy class (lazyClassName).
	Output: The source code after modifying.
1	declare a null List named lazyClassNodeList
2	for each source code file in the directory
3	for each TypeDeclaration node in this file
4	if the class name equals to lazyClassName
5	store the TypeDeclaration node into lazyClassNodeList
6	end if
7	end for
8	end for
9	
10	if the size of lazyClassNodeList is 1
11	delete the TypeDeclaration node in lazyClassNodeList
12	else
13	prompt that some classes have the same name and need to be selected by manual
14	end if

4 Experiments and Results Analysis

To evaluate accuracy and performance of the detection and elimination algorithms, four projects are used to detect the lazy classes and their brief information is listed in Table 3. Among them, SunnySport is a desktop purchase-sell-stock management system developed by Java, JHotDraw is a Java GUI framework for technical and structured graphics, the "Ice Hockey Manager" is a hockey team management game running under Linux, MacOS and Windows, and TinyUML is a free software tool for easy and quick creation of UML 2 diagrams based on Java.

TABLE 3 Brief information of the four examined projects

Measures	Sunny Sport	JHotDraw	TinyUML	IceHockey Manager
Version	1.0	5.1	0.13_02	0.3
Line of code	10265	8419	13739	18085
Number of source code files	51	144	194	218
Number of Classes/Interfaces	105	155	207	222
Number of attributes	658	331	715	1432
Number of methods	377	1314	1644	1664

Precision is used to analyse and evaluate the accuracy of the detection results. We identify the true lazy classes by manual. And the formula for calculating precision as follows:

$$\text{Precision} = \frac{TP}{TP + FP} , \qquad (4)$$

where, *TP* (True Positive) represents the number of true lazy classes in the *LCCS*. *FP* (False Positive) represents the number of false lazy classes in the *LCCS*. After examining the candidate lazy classes in *LCCS* one by one, *TP* and *FP* are obtained. Precision values of the four examined projects are listed in Table 4.

TABLE 4 Precision of the automatic detection algorithm

Project	TP	FP	TP + FP	Precision
SunnySport	2	0	2	100%
JHotDraw	6	0	6	100%
TinyUML	65	1	66	98.48%
IceHockeyManager	2	0	2	100%

In Table 4, three of the four project's precision values are 100%. All candidate lazy classes in LCCS of them are true positive instances. For example, the lazy classes in JHotDraw are DiamondFigure, NothingApplet, JavaDrawApplet, PertApplet, PatternPainter and JavaDrawViewer, to which none of other class has relationship. In TinyUML, there are 66 candidate lazy classes, including 64 test classes named XXXTest, a useless interface and a Main class, which is the entry of the project. The Main class is not a real lazy class, so it's a false positive instance. In general, the proposed approach has high accuracy for detecting lazy classes.

Moreover, we evaluate and analyse the performance of the proposed algorithm. The experiment is performed in a workstation equipped with a 2.67 GHz dual core processor and 2GB of RAM. For each project and each lazy class, we perform the detection and elimination program five times respectively. The average execution time of automatic detection is listed in Table 5.

TABLE 5 Automatic detection time of four projects

Project	Number of source code files	Line of code	Average execution time of detection (ms)
SunnySport	51	10265	1700.2
JHotDraw	144	8419	1903.4
TinyUML	194	13739	2274.4
IceHockeyManager	218	18085	2324.6

In Table 5, the average execution time is increased with the expansion of system's scale. We use the number of source code files as the *X*-axis and the average execution time of lazy classes' detection as the *Y*-axis. The linear relationship is shown in Figure 1.
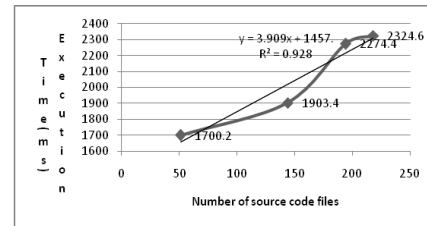
FIGURE 1 The linear relationship between number of source code files and average detection time.

Figure 1 presents the relationship between number of source code files and the execution time. In Figure 1, the thin line is a linear regression trend line. According to the algorithm analysis in Section 3, time complexity of the detection algorithm is $T(n) = O(n)$. Execution time is proportional to the number of source code files which can be used to represent the scale of system. Experimental results are in accord with the analysis results, and show that the automatic detection algorithm has good efficiency.

To evaluate performance of the automatic elimination algorithm in Table 2. We select some true lazy classes from the *LCCS*, and the average execution time is list in Table 6.

TABLE 6 Automatic elimination time of four projects

Project	Number of source code files	Class Name	Average elimination time of class (ms)	Average elimination time of project (ms)
SunnySport	51	com.SunnySport.util.StockinTableModel	218.5	218.5
		com.SunnySport.util.DButiltow	218.5	
JHotDraw	144	CH.ifa.draw.contrib.DiamondFigure	358.5	358.7
		CH.ifa.draw.samples.nothing.NothingApplet	358.5	
		CH.ifa.draw.samples.javadraw.JavaDrawViewer	359	
TinyUML	194	test.tinyuml.ui.IconLoaderTest	608.5	608.5
		org.tinyuml.model.UmlModelListener	608.5	
		test.tinyuml.draw.NullElementTest	608.5	
IceHockeyManager	218	org.icehockeymanager.ihm.clients.devgui.ihm.scenario.TMScenarioList	686.5	686.5
		org.icehockeymanager.ihm.clients.devgui.gui.icons.icons	686.5	

In Table 6, the average execution time is also increased with the expansion of system's scale. We use the same method as Figure 1 to draw the relationship diagram between the number of source code files and the average execution time. The result is shown in Figure 2.

In Figure 2, the average elimination time also has a linear relationship to the number of source code files. The experiment results are in accord with the algorithm's complexity analysis. Execution time is in direct proportion to the system's scale.

FIGURE 2 The linear relationship between number of source code files and average elimination time.

5 Conclusions

In this paper, a novel approach based on abstract syntax trees to detecting and eliminating lazy classes automatically is proposed. At the beginning, all source code files in a project transform to ASTs, then the relationships between classes are extracted from these ASTs. We analyse three kinds of classes' relationships, which are generalization, association and dependency. And we present some definitions to represent the classes set of different kinds of relationships. The candidate lazy classes set is obtained after a series of operations on these sets. We examine candidate lazy classes by manual, and remove true lazy classes finally. In order to verify our approach's correctness and evaluate its performance, four projects are used to perform the experiments for detecting and eliminating lazy classes. The experimental results show that the precision of the detection algorithms is very high. Moreover, our approach has good efficiency and it can be used for projects of different scales. Its execution time has a linear relationship to the size of system.

In the future work, we will improve and perfect our approach. On one hand, we will detect more kinds of files in a project except for source code files. Some information on classes' relationships lies in the configure files or user interface files, such as the JSP files, XML files. So, we need to detect these files in the next research work. On the other hand, we will find the isolated class groups in a system. Isolated class group is a series of classes, which have relationships to some other classes in the same group, but none of the class beyond the group needs them. All classes in an isolated class group form an islet of a set of lazy classes, and they need to be detected and eliminated together.

References

[1] Fowler M 1999 Refactoring: Improving the Design of Existing Code *Addison-Wesley Boston*

[2] Sjoberg D I, Yamashita A, Anda B C, Mockus A, and Dyba T 2013 *IEEE Transactions on Software Engineering* **39**(8) 1144-56

[3] Yamashita A, Counsell S 2013 Code smells as system-level indicators of maintainability: An Empirical Study *Journal of Systems and Software* **86**(10) 2639-53

[4] Marinescu R 2004 Detection strategies: Metrics-based rules for detecting design flaws *Proceedings 20ᵗʰ IEEE International Conference on Software Maintenance* 350-9

[5] Moha N, Gueheneuc Y G, Duchien L, Le Meur A 2010 DECOR: A method for the specification and detection of code and design smells *IEEE Transactions on Software Engineering* **36**(1) 20-36

[6] Moha N, Guéhéneuc Y G, Le Meur A F, Duchien L, Tiberghien A 2010 From a domain analysis to the specification and detection of code and design smells *Formal Aspects of Computing* **22**(3-4) 345-61

[7] Khomh F, Vaucher S, Guéhéneuc Y G, Sahraoui H 2009 A bayesian approach for the detection of code and design smells *QSIC'09 9th International Conference on* pp *Quality Software* 305-14

[8] Khomh F, Vaucher S, Guéhéneuc Y G, Sahraoui H 2011 BDTEX: A GQM-based Bayesian approach for the detection of antipatterns *Journal of Systems and Software* **84**(4) 559-72

[9] Liu H, Ma Z, Shao W, Niu Z 2012 *IEEE Transactions on Software Engineering* **38**(1) 220-35

[10] Rao A A, Reddy K N 2008 Detecting Bad Smells in Object Oriented Design Using Design Change Propagation Probability Matrix *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2008) Hong Kong China, March 2008 International Association of Engineers* 1001-7

[11] Zhang M, Hall T, Baddoo N 2011 Code bad smells: a review of current knowledge *Journal of Software Maintenance and Evolution: research and practice* **23**(3) 179-202

[12] Fontana F A, Braione P, Zanoni M 2012 Automatic detection of bad smells in code: An experimental assessment *Journal of Object Technology* **11**(2) 1-38

[13] Fontana F A, Mariani E, Morniroli A, Sormani R, Tonello A 2011 An experience report on using code smells detection tools *2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops (ICSTW)* 450-7

[14] Kuhn T, Thomann O 2013 Eclipse Corner Article Abstract Syntax Tree:
http://www.eclipse.org/articles/Article-JavaCodeManipulation_AST/index.html/ 1 Sep 2013

[15] Gamma E, Helm R, Johnson R, Vlissides J 1995 Design patterns: elements of reusable object-oriented software *Addison-Wesley: Boston*

Authors

Wei Liu, born in June, 1982, Hunan, China

Current position, grades: PhD Candidate of Computer Application Technology in Central South University, Senior Software Engineer.
University studies: PhD Student in Central South University, China, 2010.
Scientific interest: software engineering and data mining, including design patterns, refactoring, uml, reverse engineering, source code analysis and optimization, software metrics and parallel computing.
Publications: 10 publications.
Experience: more than 30 software projects.

Zhigang Hu, born in September, 1963, Shanxi, China

Current position, grades: Professor and Doctoral Supervisor at Central South University. Vice dean of the School of Software.
University studies: PhD, Mechanical Design and Theory, Central South University, China, 2003.
Scientific interest: software engineering, operating system, parallel computing, grid computing, cloud computing, embedded systems and high-performance platform.
Publications: more than 100 publications.
Experience: several research projects of National 863 Plan and the National Natural Science Foundation (NNSF).

Hongtao Liu, born in February, 1991, Hunan, China

Current position, grades: Master Degree student of Software Engineering in Central South University.
University studies: Bachelor of Engineering, Central South University, China, 2013.
Scientific interest: Software engineering, including design patterns, source code analysis, optimization and refactoring.
Publications: 5 publications.

New information multivariable optimization MGM(1,*n*) model with non equidistance and based on background value optimization

Youxin Luo[*], Qiyuan Liu, Xiaoyi Che

[1]*College of Mechanical Engineering, Hunan University of Arts and Science, Changde, 415000, P.R. China*

Abstract

The function with non-homogeneous exponential law, based on index characteristic and integral characteristic of grey model GM(1,1), was used to fit the one-time accumulated sequence, and the formula of background value was given, aiming at the problem of lower precision as well as lower adaptability in non-equidistant multivariable model MGM(1,*n*). A new information optimization model MGM(1,*n*) with non equidistance and multi variable based on background optimization was put forward, took the *m-th* component of the original sequence as initial condition, the mean relative error as objective function, and the modified one of initial value and the parameters of background value as design variables. This proposed MGM(1,*n*) model can be used in equidistance & non-equidistance modelling with higher precision as well as stronger adaptability. Examples have validated the practicability and reliability.

Keywords: multivariable, background value; optimizing, new information, non-equidistance sequence, non-equidistant MGM(1,*n*) model, least square method

1 Introduction

MGM(1,*n*) model is extended from GM(1,1) model in the case of *n* variables, and the parameters of MGM(1,*n*) model can reflect the relationships of mutual influence and restriction among multiple variables. The MGM(1,*n*) model was established [1] and the optimizing model of MGM (1,*n*) was set up by taking the first component of the sequence $\mathbf{x}^{(1)}$ as the initial condition of the grey differential equation and modifying [2]. The multivariable new information MGM(1,*n*) model taking the nth component of $\mathbf{x}^{(1)}$ as initial condition was established [3]. Take the *n*th component of $\mathbf{x}^{(1)}$ as initial condition and optimize the modified initial value and the coefficient of background value q where the form is $z_i^{(1)} = qx_i^{(1)}(k+1)+(1-q)x_i^{(1)}(k) \quad (q \in [0,1])$, and the new information MGM(1,*n*) model with multivariable was established [4]. These MGM(1,*n*) models are equidistant, and the non-equidistance multivariable MGM(1,*n*) model, with homogeneous exponent function fitting background value, was established [5]. However, non-homogeneous exponent function is more widespread, there are inherent defects in the modelling mechanism of this model. The model MGM(1,*n*) with non-equidistance and multivariable was established [6], and its background value is generated by mean value so as to bring lower accuracy. The non-equidistant and multi-variable model GM(1,n), based on non-homogeneous exponent function fitting background value, was established [7] and improves the accuracy of

the model. The constructing method for background value is a key factor affecting the prediction accuracy and the adaptability. In order to improve the accuracy of GM(1,1); some constructing methods for background value were proposed, and some non-equidistance GM(1,1) model were established [8-11]. How it is of great significance establishing non-equidistance of MGM(1,*n*) model with high precision to extend GM(1,1) model to MGM(1,*n*) model. In this study, the method constructing background value in [9] and the optimizing modelling method in [2] were absorbed, and the function with non-homogeneous exponential law was used to fit the one-time accumulated sequence. According to the new information priority principle in the grey system, A new information optimization model MGM(1,*n*) with non equidistance and multi variable based on background optimization was put forward. took the *m-th* component of the original sequence as initial condition, the mean relative error as objective function, and the modified one of initial value and the parameters of background value as design variables. This model, with higher precision, better theoretical and practical value, can be used in equidistance and non-equidistance model and extend the application range of the grey model.

---
[*]*Corresponding author* e-mail: llyx123@126.com

2 New information optimization model MGM(1,*n*) with non-equidistance and multi-variable based on background value optimization

Definition 1: Supposed the sequence $\mathbf{X}_i^{(0)} = [x_i^{(0)}(t_1),...,x_i^{(0)}(t_j),...,x_i^{(0)}(t_m)]$, if $\Delta t_j = t_j - t_{j-1} \neq const$, where $i = 1,2,...,n, j = 2,...,m$, n is the number of variables, m is the sequence number of each variable, $\mathbf{X}_i^{(0)}$ is called as non-equidistant sequence.

Definition 2: Supposed the sequence $\mathbf{X}_i^{(1)} = \{x_i^{(1)}(t_1), x_i^{(1)}(t_2),...,x_i^{(1)}(t_j),...,x_{t_j}^{(1)}(t_m)\}$, if

$x_i^{(1)}(t_1) = x_i^{(0)}(t_1)$ and $x_i^{(1)}(t_j) = x_i^{(1)}(t_{j-1}) + x_i^{(0)}(t_j) \cdot \Delta t_j$

where $j = 2,...,m$, $i = 1,2,...,n$, and $\Delta t_j = t_j - t_{j-1}$, $\mathbf{X}_i^{(1)}$ is one-time accumulated generation of non-equidistant sequence $\mathbf{X}_i^{(0)}$, and it is denoted by 1-AG0.

Suppose the original data matrix:

$$\mathbf{X}^{(0)} = \{\mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)},...,\mathbf{X}_n^{(0)}\}^{\mathbf{T}} =$$
$$\begin{bmatrix} x_1^{(0)}(t_1) & x_1^{(0)}(t_2) & ... & x_1^{(0)}(t_m) \\ x_2^{(0)}(t_1) & x_2^{(0)}(t_2) & ... & x_2^{(0)}(t_m) \\ ... & ... & ... & ... \\ x_n^{(0)}(t_1) & x_n^{(0)}(t_2) & ... & x_n^{(0)}(t_m) \end{bmatrix}, \quad (1)$$

where, $\mathbf{X}^{(0)}(t_j) = [x_1^{(0)}(t_j), x_2^{(0)}(t_j),...,x_n^{(0)}(t_j)]$

$(j = 1,2,...,m)$ is the observation value of each variable at t_j , and the sequence $[x_i^{(0)}(t_1), x_i^{(0)}(t_2),...,x_i^{(0)}(t_j),...,x_i^{(0)}(t_m)]$

$(i = 1,2,...,n, j = 1,2,...,m)$ is non-equidistant, that is, the distance $t_j - t_{j-1}$ is not constant.

In order to establish the model, firstly the original data is accumulated one time to generate a new matrix:

$$\mathbf{X}^{(1)} = \{\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \cdots, \mathbf{X}_n^{(1)}\}^{\mathbf{T}} =$$
$$\begin{bmatrix} x_1^{(1)}(t_1) & x_1^{(1)}(t_2) & \cdots & x_1^{(1)}(t_m) \\ x_2^{(1)}(t_1) & x_2^{(1)}(t_2) & \cdots & x_2^{(1)}(t_m) \\ ... & ... & ... & ... \\ x_n^{(1)}(t_1) & x_n^{(1)}(t_2) & \cdots & x_n^{(1)}(t_m) \end{bmatrix}, \quad (2)$$

where, $x^{(1)}(t_j)(j = 1,2,...,m)$ meets the conditions in the definition 2, that is,

$$x_i^{(1)}(t_j) = \begin{cases} \sum_{j=1}^{k} x_i^{(0)}(t_j)(t_j - t_{j-1}) & (k = 2,...,m) \\ x_i^{(0)}(t_1) & (k = 1) \end{cases}. \quad (3)$$

Non-equidistant and multi-variable MGM(1,*n*) model can be expressed as first-order differential equations with n variables:

$$\begin{cases} \dfrac{dx_1^{(1)}}{dt} = a_{11}x_1^{(1)} + a_{12}x_2^{(1)} + ... + a_{1n}x_n^{(1)} + b_1 \\ \dfrac{dx_2^{(1)}}{dt} = a_{21}x_1^{(1)} + a_{22}x_2^{(1)} + ... + a_{2n}x_n^{(1)} + b_2 \\ \qquad\qquad ... \\ \dfrac{dx_n^{(1)}}{dt} = a_{n1}x_1^{(1)} + a_{n2}x_2^{(1)} + ... + a_{nn}x_n^{(1)} + b_n \end{cases} \quad (4)$$

Assume $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ ... \\ b_n \end{bmatrix}$,

Equation (4) can be expressed as:

$$\frac{dX^{(1)}(t)}{dt} = AX^{(1)}(t) + B. \quad (5)$$

According to new information priority principle in the grey system, it is inadequate for utilizing new information when the first component of the sequence $\mathbf{x}_i^{(1)}(t_j)$ ($j = 1,2,\cdots,m$) is taken as the initial condition of grey differential equation. Regard the *m-th* component as the initial conditions of the grey differential equation, the continuous time response of Equation (5) is:

$$\mathbf{X}^{(1)}(t) = e^{At}\mathbf{X}^{(1)}(t_m) + \mathbf{A}^{-1}(e^{\mathbf{A}t} - \mathbf{I})\mathbf{B}. \quad (6)$$

The *m-th* component of data in Equation (6) is taken as the initial value of the solution, and then $X_i^{(0)}(t_m)$ is substituted by $X_i^{(0)}(t_m) + \beta_i$, where the dimension of $\beta = [\beta_1, \beta_2,...,\beta_n]^T$ is the same as one of $X^{(0)}(t_m)$. After restoring, the fitting value of the original data is:

$$\hat{\mathbf{X}}_i^{(0)}(t_j) = \begin{cases} \lim\limits_{\Delta t \to 0} \dfrac{\mathbf{X}_i^{(1)}(t_1) - \mathbf{X}_i^{(1)}(t_1 - \Delta t)}{\Delta t}, & j = 1 \\ (\hat{\mathbf{X}}_i^{(1)}(t_j) - \hat{\mathbf{X}}_{i t_{j-1}}^{(1)}(t_{j-1}))/(t_j - t_{j-1}), & j = 2,3,...,m \end{cases}, \quad (7)$$

where, $e^{\mathbf{A}t} = \mathbf{I} + \sum\limits_{k=1}^{\infty} \dfrac{\mathbf{A}^k}{k!} t^k$, $\mathbf{I}$ is a unit matrix.

In order to identify $\mathbf{A}$ and $\mathbf{B}$, Equation (4) is made the integration in $[t_{j-1}, t_j]$ and we can obtain:

$$x_i^{(0)}(t_j)\Delta t_j = \sum_{l=1}^{n} a_{il} \int_{t_{j-1}}^{t_j} x_i^{(1)}(t_j)dt + b_i \Delta t_j. \quad (8)$$
$$(i = 1,2,...,n; j = 2,3,...,m)$$

Note $z_i^{(1)}(t_j) = \int_{t_{j-1}}^{t_j} x_i^{(1)}(t_j)dt$, and the common formula for background value, actually based on the trapezoidal

area $z_i^{(1)}(t_j)\Delta t_j$, is appropriate when the time interval is small, that is, the change of sequence data is slow. However, when this change is sudden, the background value using the common formula often brings out the larger error, so it is more suitable for Equation (4) that parameter matrix $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ estimated by the background value in $[t_{j-1}, t_j]$ are obtained by $z_i^{(1)}(t_j) = \int_{t_{j-1}}^{t_j} x_i^{(1)}(t_j)dt$ substituting for $x_i^{(1)}(t_j)$. Based on quasi-exponentially law of the grey model and the modelling principles and methods in [9], we set that $x_i^{(1)}(t) = A_i e^{B_i t} + C_i$, where A_i, B_i, C_i are the undetermined coefficients.

Assume that the curve $x_i^{(1)}(t) = A_i e^{B_i t} + C_i$ passes through three points $(t_j, x_i^{(1)}(t_j))$, $(t_{j-1}, x_i^{(1)}(t_{j-1}))$ and $(t_{j-2}, x_i^{(1)}(t_{j-2}))$, we can obtain:

$$x_i^{(1)}(t_j) = A_i e^{B_i t_j} + C_i, \; x_i^{(1)}(t_{j+1}) = A_i e^{B_i t_{j+1}} + C_i. \quad (9)$$

Take C_i as the parameter, obtain the undetermined coefficients A_i, B_i, C_i in Equation (9):

$$A_i = \frac{[x_i^{(1)}(t_j) - C_i]^{\frac{t_{j+1}}{\Delta t_{j+1}}}}{[x_i^{(1)}(t_{j+1}) - C_i]^{\frac{t_j}{\Delta t_{j+1}}}}$$

$$B_i = \frac{\ln(x_i^{(1)}(t_{j+1}) - C_i) - \ln(x_i^{(1)}(t_j) - C_i)}{\Delta t_{j+1}}. \quad (10)$$

That Equation (10) substituting for the formula for background value $\int_{t_{j-1}}^{t_j} x_i^{(1)}(t_j)dt$ can be obtained.

$$z_i^{(1)}(t_{j+1}) = \int_{t_j}^{t_{j+1}} x_i^{(1)}dt = \int_{t_j}^{t_{j+1}} (A_i e^{B_i t} + C_i)dt =$$
$$\frac{(\Delta t_{j+1})^2 x_i^{(0)}(t_{j+1})}{\ln(x_i^{(0)}(t_{j+1}) - C_i) - \ln(x_i^{(0)}(t_j) - C_i)} + C_i \Delta t_{j+1}. \quad (11)$$

Note $\mathbf{a}_i = (a_{i1}, a_{i2}, ..., a_{in}, b_i)^{\mathbf{T}}$ $(i = 1, 2, ..., n)$ and the identified value $\hat{\mathbf{a}}_i$ of a_i can be obtained by using the least square method as follows:

$$\hat{\mathbf{a}}_i = [\hat{a}_{i1}, \hat{a}_{i2}, ..., \hat{a}_{in}, \hat{b}_i]^{\mathbf{T}} =$$
$$(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{Y}_i, i = 1, 2, ..., n, \quad (12)$$

where:

$$\mathbf{L} = \begin{bmatrix} z_1^{(1)}(t_2) & z_2^{(1)}(t_2) & ... & z_n^{(1)}(t_2) & \Delta t_2 \\ z_1^{(1)}(t_3) & z_2^{(1)}(t_3) & ... & z_n^{(1)}(t_3) & \Delta t_3 \\ ... & ... & ... & ... & ... \\ z_1^{(1)}(t_m) & z_2^{(1)}(t_m) & ... & z_n^{(1)}(t_m) & \Delta t_m \end{bmatrix}, \quad (13)$$

$$\mathbf{Y}_i = [x_i^{(0)}(t_2)\Delta t_2, x_i^{(0)}(t_3)\Delta t_3, \cdots, x_i^{(0)}(t_m)\Delta t_m]^{\mathbf{T}}. \quad (14)$$

Then the identified values of A and B can be obtained.

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} & ... & \hat{a}_{1n} \\ \hat{a}_{21} & \hat{a}_{22} & ... & \hat{a}_{2n} \\ ... & ... & ... & ... \\ \hat{a}_{n1} & \hat{a}_{n2} & ... & \hat{a}_{nn} \end{bmatrix}, \; \hat{\mathbf{B}} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ ... \\ \hat{b}_n \end{bmatrix}. \quad (15)$$

The calculated value in new information MGM(1,n) model is:

$$\hat{\mathbf{X}}_i^{(1)}(t_j) = e^{\hat{\mathbf{A}}(t_j - t_m)} \mathbf{X}_i^{(1)}(t_m) + \hat{\mathbf{A}}^{-1}(e^{\hat{\mathbf{A}}(t_j - t_m)} - \mathbf{I})\hat{\mathbf{B}}. \quad (16)$$
$$(j = 1, 2, ..., m)$$

After restoring, the fitting value of the original data is:

$$\hat{\mathbf{X}}_i^{(0)}(t_j) = \begin{cases} \lim_{\Delta t \to 0} \dfrac{\mathbf{X}_i^{(1)}(t_1) - \mathbf{X}_i^{(1)}(t_1 - \Delta t)}{\Delta t}, j = 1 \\ (\hat{\mathbf{X}}_i^{(1)}(t_j) - \hat{\mathbf{X}}_i^{(1)}(t_{j-1}))/(t_j - t_{j-1}), j = 2, 3, ..., m \end{cases} \quad (17)$$

The absolute error of the *i-th* variable is $\hat{x}_i^{(0)}(t_j) - x_i^{(0)}(t_j)$. The relative error of the *i-th* variable is: $e_i(t_j) = \dfrac{\hat{x}_i^{(0)}(t_j) - x_i^{(0)}(t_j)}{x_i^{(0)}(t_j)}100$. The mean of the relative error of the *i-th* variable is $\dfrac{1}{m}\sum_{j=1}^{m}|e_i(t_j)|$. The average error of the whole data is $f = \dfrac{1}{nm}\sum_{i=1}^{n}\left(\sum_{j=1}^{m}|e_i(t_j)|\right)$.

After taking the average error f as the objective function and β and $C = [C_1, C_2, ..., C_n]^T$ as the design variables, and the optimization function *fmincon* in Matlab 7.5 or other optimization methods [14] was used. If $C = [C_1, C_2, ..., C_n]^T$ was not optimized, $C = [0, 0, ...0]^T$, and at this time $x_{t_j}^{(1)}(i) = A_i e^{B_i t} + C_i$ is the homogeneous form as $x_{t_j}^{(1)}(i) = A_i e^{B_i t}$, and the precision of the model constructing background value is low.

3 Example

Example 1: In the calculation on contact strength, the coefficients m_a and m_b among the principal curvature function $F(\rho)$, the radius of the major axis a and the

minor b in the ellipse with the point contact is generally obtained by looking-up, and these data are extracted in Table 1 [12]:

TABLE 1 Values of $F(\rho)$, m_a and m_b

$F(\rho)$	0.9995	0.9990	0.9980	0.9970	0.9960	0.9950
m_a	23.95	18.53	14.25	12.26	11.02	10.15
m_b	0.163	0.185	0.212	0.228	0.241	0.251
$F(\rho)$	0.9880	0.9870	0.9860	0.9850	0.9840	0.9830
m_a	7.25	7.02	6.84	6.64	6.47	6.33
m_b	0.297	0.301	0.305	0.310	0.314	0.317
$F(\rho)$	0.9940	0.9930	0.9920	0.9910	0.9900	0.9890
m_a	9.46	8.92	8.47	8.10	7.76	7.49
m_b	0.260	0.268	0.275	0.281	0.287	0.292
$F(\rho)$	0.9820	0.9810	0.9800	0.9790	0.9780	0.9770
m_a	6.19	6.06	5.95	5.83	5.72	5.63
m_b	0.321	0.325	0.328	0.332	0.335	0.338

Assume m_b is as t_k, $F(\rho)$ as x_1 and m_a as x_2, non-equidistant new information optimizing MGM(1,2) model was established by using the proposed method in this study. The parameters of this model are as follows:

$$A = \begin{vmatrix} -0.0501 & 0.0003 \\ -60.1400 & 0.1394 \end{vmatrix}, \ B = \begin{vmatrix} 1.0421 \\ 76.9607 \end{vmatrix}, \ C = \begin{vmatrix} -89.9368 \\ -14.6408 \end{vmatrix},$$

$$\beta = \begin{vmatrix} 0.260702 \\ 11.7602 \end{vmatrix}.$$

The fitting value of $F(\rho)$ is:

$\hat{F}(\rho) = [0.99686, 0.99339, 0.99228, 0.9913, 0.99063, 0.9901, 0.98966, 0.98927, 0.98892, 0.98861, .98833, .98808, 0.98784, 0.98763, 0.98744, 0.98723, 0.98701, 0.98685, 0.98668, 0.98649, 0.98633, 0.98616, 0.98599, 0.98585]$

The absolute error of $F(\rho)$:

$q = [0.0026385, 0.0056107, 0.0057196, 0.0057004, 0.0053663, 0.004897, 0.0043371, 0.0037321, .0030816, 0.0023854, 0.0016663, 0.00092441, 0.00015948, -0.00062861, -0.00144, -0.0022274, -0.0030145, -0.0038487, -0.0046827, -0.0054927, -0.0063262, -0.0071595, -0.0079927, -0.0088495]$

The relative error of $F(\rho)$ (%):

$e = [-0.26398, -0.56164, -0.5731, -0.57175, -0.53879, -0.49216, -0.43632, -0.37584, -0.31065, -0.2407, -0.16832, -0.093469, -0.016142, 0.063689, 0.14604, 0.22613, 0.30635, 0.39152, 0.47685, 0.5599, .64553, 0.73131, 0.81725, 0.90578]$.

The mean of the relative error of $F(\rho)$ is 0.04822%, and one of this model is 0.49941%, therefore, this model has higher precision. In the model without optimization, the mean of the relative error of $F(\rho)$ is 0.065315%, and the one of the model is 0.79897%.

Example 2: In the conditions of the load 600N and the relative sliding speed 0.314m/s, 0.417m/s, 0.628m/s,

0.942m/s and 1.046m/s, the test data of the thin film with TiN coat are shown in Table 2 [13].

TABLE 2 Test data of the thin film with TiN coat

No.	1	2	3	4	5
Sliding speed (m/s)	0.314	0.471	0.628	0.942	1.046
Friction coefficient μ	0.251	0.258	0.265	0.273	0.288
Wear rate $\omega \times 10^{-5}$ (mg/m)	7.5	8	8.5	9.5	11

Assume sliding speed t_j, friction coefficient $X_1^{(0)}$ and wear rate $X_2^{(0)}$, non-equidistant new information optimizing MGM(1,2) model was established by using the proposed method in this study. The parameters of this model are as follows:

$$A = \begin{vmatrix} -0.1947 & 0.0102 \\ -16.2105 & 0.9536 \end{vmatrix}, \ B = \begin{vmatrix} 0.2249 \\ 4.3908 \end{vmatrix}, \ C = \begin{vmatrix} -1.1263 \\ 7.5 \end{vmatrix}$$

and $\beta = \begin{vmatrix} 0.085424 \\ 2.2022 \end{vmatrix}$.

The fitting value of $X_1^{(0)}$:

$\hat{X}_1^{(0)} = [0.25103, 0.2597, 0.265, 0.27462, 0.28471]$.

The absolute error of $X_1^{(0)}$:

$q = [2.9151e{-}005, 0.0016991, -1.607e{-}009, 0.0016164, -0.0032885]$.

The relative error of $X_1^{(0)}$ (%):

$e = [0.011614, 0.65857, -6.0642e{-}007, 0.5921, -1.1418]$.

The mean of the relative error of $X_1^{(0)}$ is 0.48082%, and one of this model is 1.3706%, thus, this model has higher precision. In the model without the optimization of β and C, the relative error mean of $X_1^{(0)}$ is 4.1851%, and the one of the model is 5.9398%. When equidistant MGM(1,3) model was used in [13], the relative error mean of $X_1^{(0)}$ is 1.6225%.

4 Conclusions

Aiming at non-equidistant multivariable sequence with mutual influence and restriction among multiple variables, the function with non-homogeneous exponential law, based on index characteristic and integral characteristic of grey model, was used to fit the one-time accumulated sequence. A new information optimization model MGM(1,n) with non equidistance and multi variable based on background optimization was put forward, took the *m-th* component of the original sequence as initial condition, the mean relative error as objective function, and the modified one of initial value and the parameters of background value as design variables. The proposed MGM(1,n) model can be used in equidistance and non-equidistance and extent the application scope of grey model. New model is with high precision and easy to use.

Examples have validated the practicability and reliability of the proposed model. This model is of important practice and theoretical significance and is worthy of promotion.

References

[1] Zhai J, Sheng J M 1997 Grey model and its application *Systems Engineering-Theory & Practice* **15**(5) 109-13

[2] Luo Y X, Li J Y 2009 Application of multi-variable optimizing grey model MGM (1,n,q,r) to the load-strain relation *The 2009 IEEE Int Conf Mechatronics and Automation (ICMA 2009)* 4023-7

[3] He Z M, Luo Y X 2009 Application of new information multi-variable grey model NMGM(1,n) to the load-strain relation *The 2009 Int. Conf. Intelligent Computation Technology and Automation (ICICTA 2009)* October 2009 **2** 11-4

[4] Luo Y X and Xiao W Y 2009 New information grey multi-variable optimization model NMGM(1,n,q,r) for the relationship of cost and variability *The 2009 Int Conf Intelligent Computation Technology and Automation (ICICTA 2009)* October 2009 **2** 120-3

[5] Wang F X 2007 Multivariable non-equidistance GM (1,m) model and its application *Systems Engineering and Electronics* **9**(3) 388-90

[6] Xiong P P, Dang Y G, Zhu H 2011 Research of modeling of multi-variable non-equidistant MGM(1,m) model *Control and Decision* **26**(1) 49-53

[7] Xiong P P, Dang Y G, Yang Y 2010 The optimization of background value in multi-variable non-equidistant model *The 19th Chinese Conf Grey Systems* 277-81

[8] Cui L Z, Liu S F, Wu Z P 2008 MGM(1,m) based on vector continued fractions theory *Systems Engineering* **26**(10) 47-51

[9] Luo Y X, Che X Y, Liu Q Y 2009 Non-equidistant GM (1,1) model with optimizing modified nth component taken as the initial value and its application to line-drawing data processing *The 2009 Int Conf Information Engineering and Computer Science (ICIECS2009)* 499-502

[10] Dai W Z, Li J F 2005 Modeling research on non-equidistance GM (1,1) model *Systems Engineering-Theory & Practice* **25**(9) 89-93

[11] Wang Z X, Dang Y G, Liu S F 2008 An optimal GM(1,1) based on the discrete function with exponential law *Systems Engineering-Theory & Practice* **28**(2) 61-7

[12] Han Z H, Dong H X 2008 Methodology and MGM(1,n) model of line chart data processing in computer aided design *Journal of Machine Design* 25(4) 18-20

[13] Luo Y X, Che X Y 2008 The grey multi-variable optimizing model and its application to analysis of the tribological behaviors of the film *Lubrication Engineering* **33**(3) 58-61

[14] Shohla M A, El-sawy A A, Nofal M, El-Zoghdy S F 2012 Using hybrid particle swarm optimization to solve machine time scheduling problem with random starting time *IJCSI International Journal of Computer Science Issues* **9**(3) 322-7

Authors

Youxin Luo, born in December, 1966, Xinhua, Hunan, China

Current position, grades: professor at the College of Mechanical Engineering, Hunan University of Arts and Science, Changde, China.
University study: B.S and M.S degrees in mechanical design & manufacturing, material engineering at Chongqing University, Huazhong University of Science and Technology, China, in 1988 and 2003 respectively.
Scientific interests: information science, grey system, mechanics, and optimizing.
Publications: 180 papers and 2 monographs.
Experience: teaching and scientific research work since 2000.

Qiyuan Liu, born in May, 1981, Changde, Hunan, China

Current position, grades: teacher at the College of Mechanical Engineering, Hunan University of Arts and Science, Changde, China.
University study: B.S and M.S degrees in mechanical manufacturing, material engineering at Xiangtan University, China, in 2002 and 2005, respectively.
Scientific interests: information science, grey system and material engineering.
Publications: 15 papers.
Experience: teaching and scientific research work since 2006.

Xiaoyi Che, born in January, 15, 1966, Changde, Hunan, China

Current position, grades: professor at the College of Mechanical Engineering, Hunan University of Arts and Science, Changde, China.
University study: B.S and M.S degrees in mechanical manufacturing, material engineering at Shengyan Institute of Technology, Huazhong University of Science and Technology, China, in 1988 and 2003, respectively.
Scientific interest: Information science, grey system, CNC and optimum design.
Publications: 30 papers.
Experience: teaching and scientific research work since 1988.

The study on elliptical flange hole forming based on finite element analysis

Jiansheng Xia[1, 2*], Shasha Dou[1, 2]

[1]*Yancheng Institute Of Technology, Yancheng City, Jiangsu Province, China, 224051*

[2]*Nanjing University of Aeronautics & Astronautics, Nanjing City, Jiangsu Province, China, 211106*

Abstract

The flange hole forming is a complex process, Under the assumption of Prandtl-Reuss flow rule and von Mises yield criterion, the incremental elasto-plastic large deformation finite element model was established based on the Updated Lagrangian Formulation (ULF).The elasto-plastic conversions of boundary and deformation are reduced with r-min rule. The friction phenomenon of slippage and viscosity at the boundary interface is revised with increment of revision Coulomb rule. The increment rules are led into the whole stiffness matrix, and derived out the stiffness equation. The studies show that the influence on steel elliptical hole flange forming deformation is influenced by punch structure and parameter. The dates show that finite element simulation and experimental result have a good consistency.

Keywords: elasto-plastic, FEM Simulation, elliptical hole flange

1 Introduction

Sheet metal forming is a common material processing method, which can be divided into stretching, bending, drawing and flanging, etc. [1]. Its parts can be used in automotive, stationery manufacturing, household appliances industry, when provided with the support group or for pipeline connection and other purposes.

The manufacturing process is using the elliptical punch to stretch forming the sheet, and the blank will bend along the punch radius to form an elliptical shape hole. The hole expands with the dropping of the punch, and the thickness near the hole will gradually thinning, and result the neck and cracked phenomenon. Therefore, if using the finite element to analysis the flange forming process, instead of the trial and error method which are used in general mold factory, it will reduce the costs and shorten development time away [2].

Tang [3] studied the different punches to analyse the distribution of stress and strain in the flange forming with the shell theory and ignoring the bending effect.

Huang and Chen [4,5] investigated the flange hole shape with the different punch radius and shapes. The results show there is the linear relationship between the initial diameter and after stretching.

Takuda and Hatta [6,7] used the rigid-plastic finite element to simulate the sheet metal forming and used ductile fracture criterion to predict zirconium sheet stretch forming limit. The results show that the extension of zirconium sheet is high, but the stretch is low.

Leu [8] studied the numerical analysis and experiment of flange process with the incremental elastic-plastic finite element method. The results can effectively predict the forming process: when hardening index and orthogonal anisotropy increased, the maximum hole expansion ratio also increased.

Huang and Chien [9] studied the forming process of the flange hole with the frusto-conical punches and different radius, and found the frusto-conical taper punch radius does not affect the forming, the maximum load decreases with the punch radius increasing.

In this paper, the steel sheet SPCC is analysed with the finite element method, some relationships are studied, such us: relationship between punch load and displacement, distribution of stress and strain, distribution of thickness, and verify by actual experiments. It used to reference for operation process and altered design.

2 Fundamental theory

2.1 VIRTUAL WORK PRINCIPLE

It descripts the elastic-plastic deformation with the updated Lagrangian formulation ULF [6], the Virtual work principle formulation can be shown as follows:

$$\int_{V^E}\left(\overset{\triangledown}{\sigma}_{ij} - e\sigma_{ik}\dot{\varepsilon}_{kj}\right)\delta\dot{\varepsilon}_{ij}dV + \int_{V^E}\sigma_{jk}L_{ik}\delta L_{ij}dV = \int_{S_f}\dot{f}\delta v_i dS \text{ ,(1)}$$

where, $\overset{\triangledown}{\sigma}_{ij}$ is the Cauchy stress tensor, $\dot{\varepsilon}_{kj}$ is the rate of stress tensor, $\dot{\varepsilon}_{ij}$ is the strain tensor, σ_{jk} is the rate of strain tensor, $\delta\dot{\varepsilon}_{ij}$ is the virtual strain tensor of the point, δL_{ij} is the virtual velocity gradient tensor of the point, δv_i is the

Xia Jiansheng, Dou Shasha

velocity component, $\dot{f}$ is surface force component, L_{ij} is velocity gradient tensor, V is unit volume, S is unit surface area.

2.2 CONSTITUTIVE RELATION

In preparing the theory of elasto-plasticity, we have made certain assumptions[10]:

    1) The material is homogeneous and isotropic;

    2) There is no strain before manufacturing;

    3) Temperature effect don't consider when manufacturing;

    4) It obeys the laws of the Hooke's Law in elastic stage;

    5) It obeys the von Mises yield rule and Prandtl-Reuss plastic flow rule;

    6) It contains Isotropic strain hardening in constitutive equation;

    7) There are elastic strain stage and plastic strain stage in material strain rate;

    8) Punch, die and holder are steel structure;

    9) The Bauschinger effect don't consider in reverse unloading.

    Taking into account mentioned above, the constitutive relation can be written as follows:

$$\overset{\circ}{\sigma}_{ij} = C^{ep}_{ijmn}\dot{\varepsilon}_{mn}, \tag{2}$$

$$C^{ep}_{ijmn} = C^{e}_{ijmn} - \frac{C^{e}_{ijkl}C^{e}_{uv}\dfrac{\partial f}{\partial \sigma_{kl}}\dfrac{\partial f}{\partial \sigma_{uv}}}{C^{e}_{kluv}\dfrac{\partial f}{\partial \sigma_{kl}}\dfrac{\partial f}{\partial \sigma_{uv}} + H'\dfrac{\sigma_{uv}\dfrac{\partial f}{\partial \sigma_{uv}}}{\bar{\sigma}}}, \tag{3}$$

where: $\overset{\circ}{\sigma}_{ij}$ is Jaumann differential of σ_{ij}, C^{ep}_{ijmn} is the elastic-plastic module, C^{e}_{ijmn} is Elastic module, f is the initial yield function, H' is the strain hardening rate, $\bar{\sigma}$ is Von Mises yield function, so the Matrix form of C^{ep}_{ijmn} can be expressed as below:

$$[C^{ep}] = [C^{e}] -$$

$$\frac{1}{S}\begin{bmatrix} S_1^2 & S_1S_2 & S_1S_3 & S_1S_4 & S_1S_5 & S_1S_6 \\ & S_2^2 & S_2S_3 & S_2S_4 & S_2S_5 & S_2S_6 \\ & & S_3^2 & S_3S_4 & S_3S_5 & S_3S_6 \\ & & & S_4^2 & S_4S_5 & S_4S_5 \\ & & & & S_5^2 & S_5S_6 \\ symm & & & & & S_6^2 \end{bmatrix}, \tag{4}$$

where,

$$S = \frac{4}{9}\bar{\sigma}^2 H' + S_1\sigma'_{xx} + S_2\sigma'_{yy} + S_3\sigma'_{zz} +$$
$$2S_4\sigma'_{yz} + 2S_5\sigma'_{zx} + 2S_6\sigma'_{xy}, \tag{5}$$

$$S_1 = 2G\sigma'_{xx}, \; S_2 = 2G\sigma'_{yy} \; S_3 = 2G\sigma'_{zz}, \tag{6}$$

$$S_4 = 2G\sigma'_{yz}, \; S_5 = 2G\sigma'_{zx}, \; S_6 = 2G\sigma'_{xy}, \tag{7}$$

where σ'_{ij} is deviator of σ_{ij}, G is the friction flow potential, $G = \sigma_1^2 + \sigma_2^2$, $[C^e]$ is the equation in minimum strain, which can be expressed as below:

$$[C^e] = \frac{E}{1+v}\begin{bmatrix} \dfrac{1-v}{1-2v} & \dfrac{1-v}{1-2v} & \dfrac{1-v}{1-2v} & 0 & 0 & 0 \\ & \dfrac{1-v}{1-2v} & \dfrac{1-v}{1-2v} & 0 & 0 & 0 \\ & & \dfrac{1-v}{1-2v} & 0 & 0 & 0 \\ & & & \dfrac{1}{2} & 0 & 0 \\ & & & & \dfrac{1}{2} & 0 \\ symm & & & & & \dfrac{1}{2} \end{bmatrix}, \tag{8}$$

where, E is modulus of elasticity, v is Poisson's ratio. If the material is homogeneous and isotropic, the Elasto-plastic rate equation can be written:

$$\overset{\circ}{\sigma}_{ij} = \frac{E}{1+v}\left[\delta_{ik}\delta_{jl} + \frac{v}{1-2v}\delta_{ij}\delta_{kl} - \frac{3\alpha\left(\dfrac{E}{1+v}\right)\sigma'_{ij}\sigma'_{kl}}{2\bar{\sigma}^2\left(\dfrac{2}{3}H' + \dfrac{E}{1+v}\right)}\right]\dot{\varepsilon}_{kl}. \tag{9}$$

When $\alpha = 1$, it is a plastic stage; when $\alpha = 0$, it is an elastic stage or unloading stage.

Equivalent stress and equivalent plastic strain relations can express by n-power law equation:

$$\dot{\sigma} = C\left(\varepsilon_0 + \dot{\varepsilon}_p\right)^n, \tag{10}$$

where: C is material constant, n is strain hardening index; $\dot{\sigma}$ is the equivalent stress, $\dot{\varepsilon}_p$ is the equivalent plastic strain, ε_0 is the initial strain.

2.3 THE FINITE ELEMENT FORMULA

Finite element analysis is the method that the structure is divided into many small units called discrete entity. Based on Large deformation stress and stress rate relation, the finite deformation of Update Lagrangian Formulation, material constitution relationship, the velocity distribution of each unit is shown below:

$$\{v\} = \leq [N]\{\dot{d}\}, \tag{11}$$

$$\{\dot{\varepsilon}\} = \leq [B]\{\dot{d}\}, \tag{12}$$

$$\{L\} = \leq [M]\{\dot{d}\} , \tag{13}$$

$$[K]\{\Delta u\} = \{\Delta F\} , \tag{14}$$

where $[N]$ is shape function, $\{\dot{d}\}$ is nodal velocity, $[B]$ is strain rate-velocity matrix, $[M]$ is velocity gradient-velocity matrix.

where:

$$[K] = \sum_{\langle E \rangle} \int_{V^e} [B]^T \left([C^{ep}] - [Q] \right) [B] dV + \sum_{\langle E \rangle} \int_{V^e} [E]^T , \tag{15}$$

The principle of virtual work equation and the constitutive equation based on update Lagrangian is linear equation. The equation can be written by the form of incremental representation.

After finite element discrimination, the large deformation rigid general equation is written as below:

$$\{\Delta F\} = \sum_{\langle E \rangle} \int_{S^E} [N]^T \{\dot{f}\} dS \Delta t , \tag{16}$$

$[K]$ is the overall Elasto-plastic stiffness matrix, $\{\Delta F\}$ is the nodal displacement increment, $\{\Delta u\}$ is the nodal forces incremental, $[Q]$ and $[Z]$ are stress correction matrix.

$$Q = \begin{bmatrix} 2\sigma_{xx} & 0 & 0 & \sigma_{xy} & 0 & \sigma_{xz} \\ & 2\sigma_{yy} & 0 & \sigma_{xy} & \sigma_{zy} & 0 \\ & & 2\sigma_{zz} & 0 & \sigma_{zy} & \sigma_{xz} \\ & & & \frac{1}{2}(\sigma_{xx} + \sigma_{yy}) & \frac{1}{2}(\sigma_{zx}) & \frac{1}{2}(\sigma_{zy}) \\ & & & & \frac{1}{2}(\sigma_{yy} + \sigma_{zz}) & \frac{1}{2}(\sigma_{xy}) \\ symm & & & & & \frac{1}{2}(\sigma_{xx} + \sigma_{zz}) \end{bmatrix} , \tag{17}$$

$$Z = \begin{bmatrix} \sigma_{xx} & 0 & 0 & \sigma_{xy} & \sigma_{xz} & 0 & 0 & 0 & 0 \\ & \sigma_{yy} & 0 & 0 & 0 & \sigma_{xy} & \sigma_{yz} & 0 & 0 \\ & & \sigma_{zz} & 0 & 0 & 0 & 0 & \sigma_{xz} & \sigma_{yz} \\ & & & \sigma_{yy} & \sigma_{yz} & 0 & 0 & 0 & 0 \\ & & & & \sigma_{zz} & 0 & 0 & 0 & 0 \\ & & & & & \sigma_{xx} & \sigma_{xz} & 0 & 0 \\ & & & & & & \sigma_{zz} & 0 & 0 \\ & & & & & & & \sigma_{xx} & \sigma_{xy} \\ symm & & & & & & & & \sigma_{yy} \end{bmatrix} . \tag{18}$$

2.4 FRICTION PROCESSING

There is friction in sheet forming process, so we need to pay attention to materials and tools of the interface conditions [11]. When the material moves along the tool surface curve of the slide, the contact force can be expressed as:

$$F = F_l l + F_n n , \tag{19}$$

where, F_l is radial force and F_n is normal force, and differential equation of F can be expressed as:

$$\dot{F} = \dot{F}_l l + F_l \dot{l} + \dot{F}_n n + F_n \dot{n} , \tag{20}$$

where, differentials of $\dot{l}$ and $\dot{n}$ are expressed as:

$$\dot{l} = -\Delta u_l^{rel} / R , \tag{21}$$

$$\dot{n} = \Delta u_l^{rel} / R , \tag{22}$$

where, R is tool radius, Δu_l^{rel} is the local relative velocity between the tool and node, and the nodes relative speed can be expressed as:

$$\Delta u_l^{rel} = \Delta u_l - \dot{u}_{tool} \sin \theta , \tag{23}$$

where, Δu_l is the contact tangent displacement increment of nodes, $\dot{u}_{tool}$ is the displacement increment of tooling, θ is the rotation angle.

The increment formula of $\dot{F}$ is expressed as follow:

$$\dot{F} = \left(\dot{F}_l - F_n \Delta u_l / R + F_n \dot{u}_{tool} \sin \theta / R \right) \cdot l +$$
$$\left(\dot{F}_n - F_l \Delta u_l / R - F_l \dot{u}_{tool} \sin \theta / R \right) \cdot n. \tag{24}$$

Rigid matrix governing equation of the contact nodes is expressed below:

$$\begin{bmatrix} K & ... & \\ ... & K_{11} + F_n / R & K_{12} \\ ... & K_{21} + F_n / R & K_{22} \end{bmatrix} \begin{Bmatrix} ... \\ \Delta u_l \\ \Delta u_n \end{Bmatrix} = \begin{Bmatrix} ... \\ \dot{F}_l + F_n \dot{u}_{tool} \sin \theta / R \\ \dot{F}_l - F_n \dot{u}_{tool} \sin \theta / R \end{Bmatrix}. \tag{25}$$

2.5 INCREMENTAL STEPS OF R_{MIN} METHOD

Using the elastic plastic finite element method with large deformation method, also called the Yamada r_{min} method. Each incremental step value is equal to incremental displacement of initial deformation increment of the tooling. Adopting the method of updated Lagrangian formulation, calculating each increment of displacement, strain, stress, load, springback value after forming the final shape of sheet mental in unloading condition, the value of load incremental in each step is controlled by r_{min} formula, which is shown as below:

$$r_{min} = MIN(r_1, r_2, r_3, r_4, r_5) , \tag{26}$$

where, r_1 is The maximum allowable strain increment, r_2 is the maximum allowable rotation increment, r_3 is the minimum value in all elastic elements, r_4 is contact position between free node and tooling, r_5 is discontent position between free node and tooling.

3 Numerical analysis flow

Based on the finite deformation theory, ULF equation and r_{min} method, a set of effective analysis of sheet metal forming process is established. Firstly, a 3d part and mold is designed with the NX software, and then mesh them with NASTRAN software. Secondly, the meshed models are drawn into the data file and did finite element analysis. The simulation flow chart is shown in Figure 1.
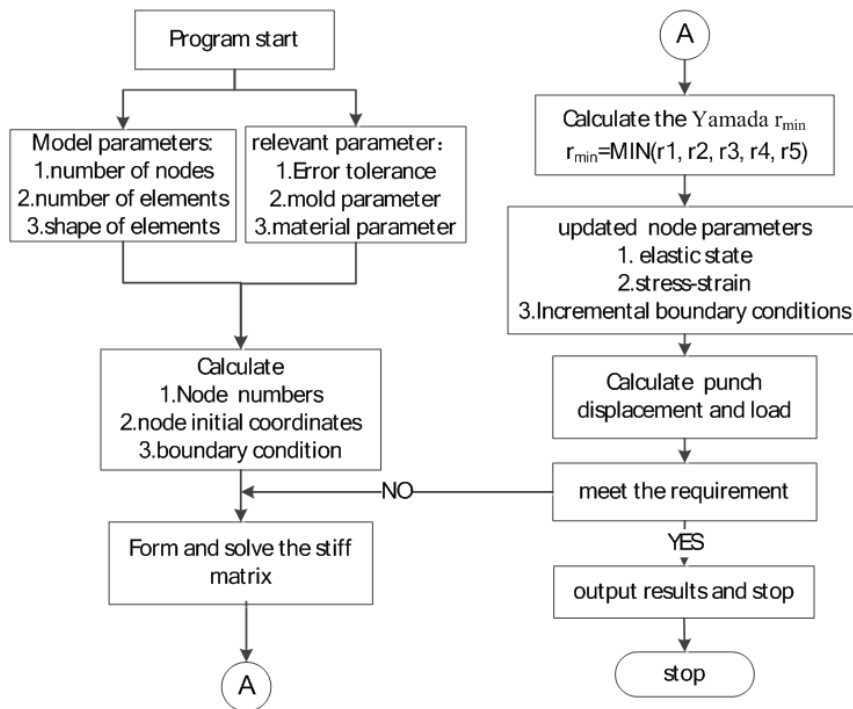
FIGURE 1 Numerical simulation of flow chart

Based on the theory upwards, the research of steel elliptical cup drawing is studied, including relationship between the punch load and displacements, stress and strain, thickness, spring-back and warpage. Simulation experimental parameters were carried out, which are friction coefficient (μ), punch radius (r_p), die radius (R_d).

The parameters of warpage problems are verified by the experiment are optimized and served a reference for drawing designer.

The whole structure is composed of punch, die and blank holder. The model picture was shown as Figure 2.
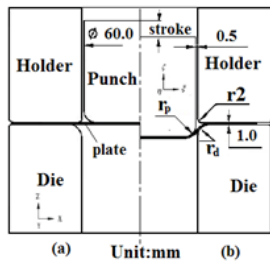
FIGURE 2 Sheet metal and die size chart; (a) before deformation; (b) after deformation

The initial relation of part and die is shown in Figure 2a, also, the punch down a certain travel case is shown Figure 2b. It takes two coordinates to solve the problem, which are fixed coordinates (X, Y, Z) and local coordinates (ξ, η, ζ). It uses the fixed coordinates (X, Y, Z) when nodes do not contact with the tool, and uses the local coordinates (ξ, η, ζ) when nodes contact with the tooling. Using coordinates rule based on the right-hand rule. L-axis is the tangential direction of contact line between the part and tools when n-axis is the normal direction.

The contact condition of each node of plates will change depending on deformation in sheet metal forming. When the displacement increment is zero, the boundary conditions of increment displacement of the next node will changes to free node boundary conditions. When sheet contacts the tools, contact condition is changed to the boundary conditions, which bases on the generalized r_{min} method.

Blanks preparation: JIS SPCC steel sheet, cutting into the outer diameter 130.0mm and an initial elliptical hole in the centre of the sheet by CNC, the long axis of elliptical size 23mm, short axis dimension 12mm.

Experimental arrangement: the sheet metals are put on hydro forming machine, the centre of the sheet is consistent with the mold. Set pressure of the pressure 160kN, the punch speed 1.0mm/s. Measured the experimental data of header punch load and the stroke of the punch. Finally, measured and recorded the hole of the flange height was measured with callipers.

JIS SPCC material stainless steels are provided by a china steel Crop, of which the mechanical properties as shown in Table1 as below [12].

TABLE 1 Mechanical properties of JIS SPCC

stress-strain relationship: $\bar{\sigma}=563.64(0.011+\bar{\tau}_p)^{0.2781}$	
initial thickness: $t = 0.6$mm	Poisson's ratio: $\upsilon=0.3$
yield stress: $\sigma y=163.5$MPa	Anisotropy value: $r_0 =1.71$, $r_{45} = 1.52$, $r_{90} = 2.11$
Yang coefficient: $E=2.1\times10^5$ MPa	

Because of the symmetrical sheet model, 1/4 model is taken to analysis.

It uses the quadrilateral segmentation of degenerated shell element in sheet metal meshing, when the die meshing uses the triangle segmentation.

4 Results

4.1 DISTRIBUTION OF STRESS AND STRAIN

The hole elliptical flange stress distribution as shown in Figure 3. As can be seen from the figure, the stress near the long axis of the hole is maximum, the maximum value is 395MPa, because the holes of the flange forming influences by tensile stress, and the curvature is large near the long axis of the hole, the circumferential stress has concentration phenomenon. Stress near the short axis of the holes gradually decreases in the direction of elliptical, which is due to the short axis of the inner peripheral edge of the blanks curvature of the hole is small, the stress in this area is small, in addition, the stress value is less than the value of stress at the long axis.
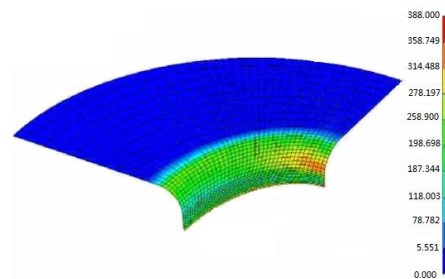
FIGURE 3 The figure of stress distribution

4.2 FLANGE HEIGHT ANALYSIS

The compare of heights of elliptical hole flange between numerical analysis and experimental is shown in Figure 4, the measured points from the long axis to short axis along the axis elliptical. As can be seen from the figure, the maximum height position at the short axis. The value is compared with the experimental results, the error is less than 2%.
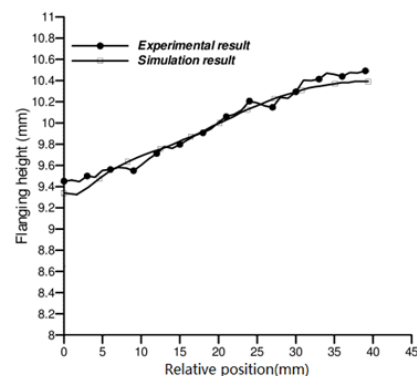
FIGURE 4 The flange height distribution

4.3 THICKNESS CHANGE ANALYSIS

The sheet mental is used to do drawing experiment analysis. The thickness data are measure along the direction of long-axis and short-axis and as shown in Figure 5. As can be seen from the Figure 5a.
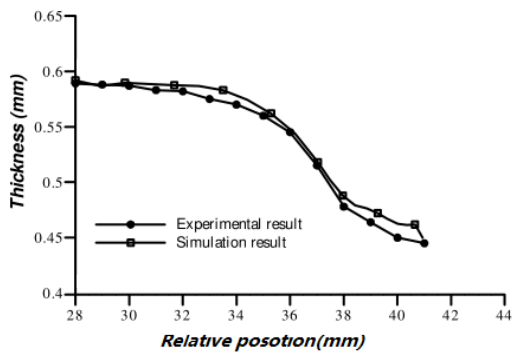
a) The thickness in long-axis
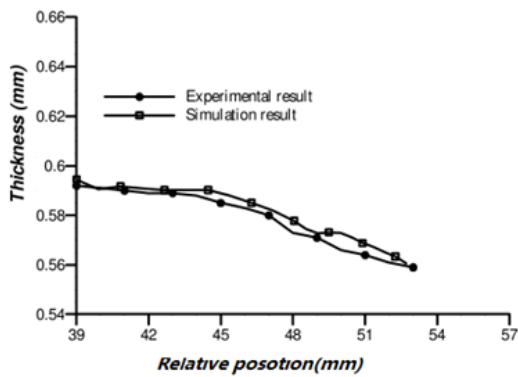
b) The thickness in short-axis

FIGURE 5Thickness change analysis in long axis and short axis

Because the part was held by holder and die in this area, the thickness changed a little from the area 0.0mm to 28.0mm. After that area, the thickness is thin gradually, until to the minimum value. The reason is that the sheet is affected by the maximum tensile stress, and the hole becomes larger and the thickness become thin. As can be seen from the Figure 5b), because the part was held by holder and die in this area, the thickness changes a little from the area 0.0mm to 39.0mm. After that area, the thickness is thin gradually, until to the minimum value. However, because the short-axis of curvature and Tensile stress are small, the value also changes a little.

The thinnest thickness is in the bottom area near the long-axis, which is the 0.45mm and slightly larger than the rapture thickness, then the part can be finished.

4.4 FORMING LIMIT ANALYSIS

In order to find out the limit thickness of elliptical flange, some experiments are carried out with the different diameters of sheet, and set the value 0.41mm as the criterion of rupture. If the thickness is thinner, the program will judge that the sheet is in the limit thickness. FR (Forming ratio) LFR (Limit Forming Ratio) and EFR (Excessive Forming Ratio) are referred by Huang and Chien [9].

DR (Drawing Ratio), LDR (Limit Drawing Ratio) and EDR (Excessive Drawing Ratio) are used to calculate and judge, which are defined by Huang [9].

$$FR = \frac{D_p}{D_h} = \frac{C_p}{C_h}, \tag{27}$$

$$LFR = \frac{D_P}{D_{h,min}} = \frac{C_P}{C_{h,min}}, \tag{28}$$

$$EDR = \frac{D_p}{D_{h,f}} = \frac{C_p}{C_{h,f}}, \tag{29}$$

where, D_p is the diameter of punch; D_h is the diameter of sheet with necking or fracture; D_h, min is the maximum diameter of sheet without necking or fracture; D_p is the perimeter of the elliptical punch; $D_{h,f}$ is the perimeter of the sheet without necking or fracture; C_p is the circumference of punch, C_h is the circumference of initial inner elliptical hole with necking or fracture; $C_{h,min}$ is the maximum the circumference of initial inner elliptical hole without necking or fracture; $C_{h,f}$ is circumference of initial inner elliptical hole with part necking or fracture. The experimental arrangement is shown in Table 2 and the simulation results are shown as Figure 6. When a = 23mm b = 12mm (FR=1.38) and a = 22mm b = 11mm (LFR=1.46), the forming complete without necking or rupture. When a = 21mm b = 10mm (EFR=1.55), the forming with necking or rupture. So the LFR=1.46 is the limit value.

The other experiments were carried out and found the result is closed to the simulation value.
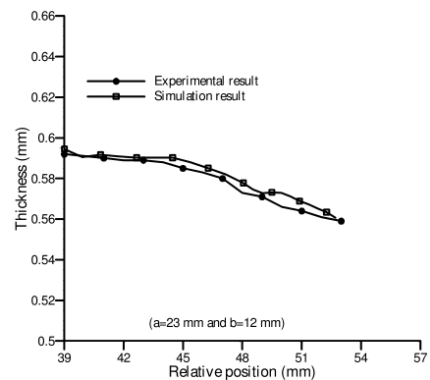
a) thickness distribution when FR=1.38
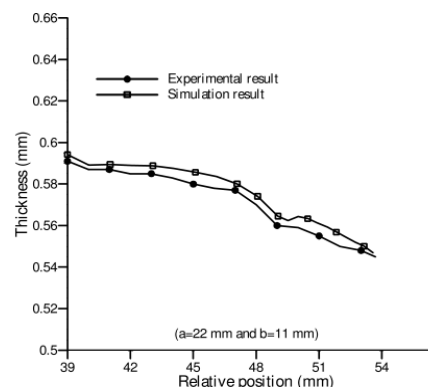
b) thickness distribution when FR=1.46
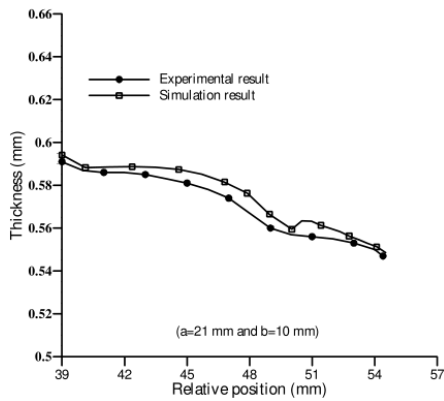
c) thickness distribution when FR=1.38

FIGURE 6 The distribution of thickness

4.5 PUNCH RADIUS ANALYSIS

In order to study the relationship between the initial inner elliptical hole and punch or punch radius, five experiments are carried out. The radius of punch and punch fillet are 3.0mm, 5.0mm, 7.0mm, 9.0mm and 11.0mm, respectively. The results are as shown in Table 2 as below.

As can be seen from the Figure 7a, maximum punch load decreases with increasing of punch radius and initial elliptical hole diameter.

As can be seen from the Figure 7b, maximum punch stroke increases with increasing of punch radius and initial elliptical hole diameter.

TABLE 2 Maximum punch load and stroke with the different radius of punch

Elliptical sizes	A=23mm, b=12mm		A=22mm, b=11mm		A=21mm, b=10mm		A=20mm, b=9mm	
Pun radius	Load(N)	Stroke(mm)	Load(N)	Stroke(mm)	Load(N)	Stroke(mm)	Load(N)	Stroke(mm)
3.0mm	5500	13.32	6680	14.08	7810	14.78	9200	15.23
5.0mm	4865	15.21	5980	16.02	6875	16.89	8000	17.80
7.0mm	4512	17.05	5285	18.01	6476	18.56	7420	19.55
9.0mm	4000	19.18	4850	20.26	5753	20.96	6730	21.89
11.0mm	3770	21.21	4505	22.08	5208	23.05	6210	24.00

a) Load distribution

5 Conclusion

Based on the numerical analysis and experimental results, combined with finite element method with the incremental Elasto-plastic theory, analysed Stress distribution, thickness analysis, forming limit analysis and punch radius analysis. It obtains the following conclusions:

1) Maximum stress and minimum thickness values occur in the area of the long-axis of the hole. This is due to the hole withstand greater the tensile stress near the long-axis area, and the tensile stress will make thickness thinner. The thickness changed small in the short-axis, because the curvature and stress are small.

2) In the long-axis, maximum punch load decreases with increasing punch radius and initial elliptical hole diameter.

3) Maximum punch stroke increases with increasing punch radius and decreasing of initial elliptical hole diameter.

4) As the forming limit defined by the inner circumference and diameter of elliptical punch, the experimental flange forming limit ratio of 1.46.

b) Stroke distribution

FIGURE 7 Load and stroke distribution with the radius of punch

References

[1] Xia J S Dou S S 2013 Experimental Research on the Flanging Height Adjustment Based on Environmental Way *International Journal of Applied Environmental Sciences* **8**(20) 2470-89

[2] Huang Y M, Chen T C 2005 Influence of Blank Profile on the V-Die Bending Camber Process of Sheet Metal *International Journal of Advanced Manufacturing Technology* **25**(7) 668-77

[3] Tang S C 1981 Large Elastic-Plastic Strain Analysis of Flanged Hole Forming *Computers & Structures* **13**(1-3) 363-70

[4] Kumagai T, Saiki H, Meng Y 1999 Hole Flanging with Ironing of Two-Ply Thick Sheet Metals *Journal of Materials Processing Technology* **89-90** 51-7

[5] Huang Y M, Chen T C 2001 The Formability Limitation of the Hole-Flanging Process *Journal of Materials Processing Technology* **117**(1) 43-51

[6] Huang Y M, Chien K H 2001 Influence of the Punch Profile on the Limitation of Formability in Hole-Flanging Proces *Journal of Materials Processing Technology* **113**(1) 720-4

[7] Takuda H, Hatta N 1998 Numerical Analysis of Formability of a Commercially Pure Zirconium Sheet in Some Forming Processes *Materials Science and Engineering* **242**(1) 15-21

[8] Leu D K 1996 Finite-Element Simulation of Hole-Flanging Process of Circular Sheets of Anisotropic Materials *International Journal of Mechanical Sciences* **38**(8) 917-33

[9] Huang Y M, Chien K H 2002 Influence of Cone Semi-Angle on theFormability Limitation of the Hole-Flanging Process *International Journal of Advanced Manufacturing Technology* **19**(8) 597-606

[10] Sousa L C, Castro C F, António C A C 2006 Optimal Design of V and U Bending Processes Using Genetic Algorithms *Journal of Material Processing Technology* **172**(1) 35-41

[11] Liu G, Lin Z, Xu W, Bao Y 2002 Variable Blankholder Force in U-Shaped Part Forming for Elimination Springback Error *Journal of Material Processing Technology* **120**(1-3) 259-64

[12] Samuel M 2000 Experimental and numerical prediction of springback and side wall curl in u-bending of anisotropic sheet metals *Journal of Materials Processing Technology* **3**(105) 382-93

Authors

Jiansheng Xia, born in September, 1980, Yancheng City, Jiangsu Province, P.R. China

Current position, grades: lecturer at Yancheng Institute of Technology, China.
University studies: BSc in Mechanical engineering and automation at Northeast Agricultural University in China. MSc at Northeast Agricultural University.
Scientific interests: mold design, CAD/CAE/CAM technologies.
Publications: more than 10 papers.
Experience: teaching experience of 8 years, 2 scientific research projects.
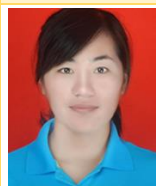
Shasha Dou, born in December, 1982, Yancheng City, Jiangsu Province, P.R. China

Current position, grades: lecturer of Yancheng Institute of Technology, China.
University studies: BSc in mechanical engineering and automation at Shandong University of Technology in China, MSc at Shandong University of Technology.
Scientific interests: plastic mold design, CAD/CAE/CAM technologies.
Publications: more than 10 papers.
Experience: teaching experience of 8 years, 1 scientific research projects.

Adaptive IMC for variable parameter systems with large dead-times

Lianhua Hu[1, 2]*, Xinping Li[2], Wei Tang[1]

[1]*School of Shaanxi University of Science & Technology, Xi'an City, Shaanxi Province, China, 710021*

[2]*Shaanxi Key Laboratory on Paper Technology and Specialty Papers, Xi'an City, Shaanxi Province, China 710021*

Abstract

Dead-time variation can cause object mismatches in traditional internal model control (IMC) systems and may result in a significant overshoot. The adjustment time may increase due to this variation and sometimes even causes oscillation instability. An adaptive IMC method is proposed in this paper to solve the problem of variable parameters in the control process. The adaptive law is designed to optimize local parameters relative to the output error of both the plant and model, ensuring that the model approximates the real plant. The control structure adopts IMC and simulation results show that this type of control structure exhibits some promising characteristics, such as high accuracy, robustness and disturbance rejection. This model is therefore suitable for systems with large dead-time varying parameters.

Keywords: Large dead-time, Internal model control (IMC), Adaptive, Variable parameter

1 Introduction

The internal model control (IMC) algorithm is widely used in dead-time process industries because of its simplicity and practical success. A well-designed IMC controller has been proved to be sufficient for a large number of dead-time control loops. However, when dead-time variation is present, although advanced control techniques can provide significant improvements, in general the output of a conventional IMC cannot adapt quickly enough to reflect the current system conditions and results in a significant overshoot. Simultaneously, the adjustment time increases and sometimes even causes oscillation instability. Where the dead-time is varying and the object parameters are also changing, the problem becomes even more complicated, and a quality control system cannot be achieved using conventional Smith predictor, optimal control or other similar algorithms [1-3]. With ordinary dead-time systems, overcoming the impact of dead-time is a key issue in the design process of the controller. For a dead-time system where the object parameters are varying, we not only need to overcome the effects due to time delays, but also need to identify changes to the object parameters. This paper presents a design for an adaptive identification IMC to track changes in object parameters and overcome the impact of dead-time. Additionally, this design benefits from IMC's characteristics such as simple structure, easy and intuitive design, less online parameter adjustment, and a simple and clear adjustment method that does not exhibit any static error when in steady-state. It also efficiently improves the anti-jamming performance and large dead-time system control effect [4-7].

2 Adaptive IMC Design

IMC was derived by Brosilow and Tong in 1978 on the basis of Smith Predictor compensation control. The model can be represented as a basic control structure using a Single Input/Single Output (SISO) block diagram. Using the general delay system IMC design presented in [4], we have selected a first-order system with a time delay controlled object. The control structure is shown in figure1.

The object model is:

$$G_M(s) = [K_m/(T_m s+1)]e^{-\tau_m s}. \tag{1}$$

The controller is:

$$G_c(s) = (T_m s+1)/K_m. \tag{2}$$
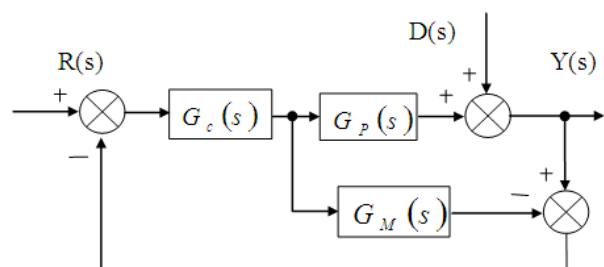
FIGURE 1 IMC structure

* *Corresponding author* e-mail: hulianhua@sust.edu.cn

The model can be considered to be a PD controller, since noise is inevitable in real-world industrial processes; therefore we have added a first-order filter after the controller to improve the robustness of the system. The larger the time constant of the filter, the better the system robustness, however the transition process becomes longer as the time constant increases [1]. The control system has good quality control once the model matches the object, but when there is a mismatch, the quality of the system will deteriorate, or even become unstable. An adaptive mechanism is therefore important to identify object parameters, thus improving the robustness of the system. We will now examine the adaptive identification object parameters in more detail.

Suppose the transfer function of the actual object is:

$$G_p(s) = K\frac{N(s)}{D(s)}e^{-\tau_P s}. \tag{3}$$

Due to the complexity of the industrial process, the parameters within either of the two transfer functions are likely to change.

The parameters can be divided into two categories: (a) rational fraction polynomial parameters, such as the Molecular denominator polynomial coefficients of $K\frac{N(s)}{D(s)}$, and (b) time delay parameters. Further detail on identification of the rational fraction polynomial parameters is given in [3], in this paper, these parameters are fixed as constants. This work focuses on delay parameter identification with local parameter optimization theory to design the adaptive law. The adaptive identification delay parameter diagram is shown in figure 2.
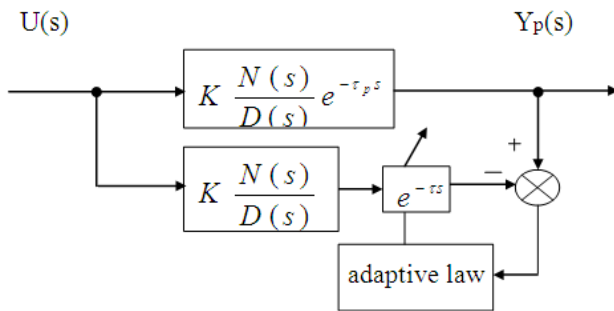
FIGURE 2 Adaptive identification agency

Assuming the prediction model is:

$$G_m(s) = K\frac{N(s)}{D(s)}e^{-\tau s}. \tag{4}$$

The selected performance indicators are:

$$J = \frac{1}{2}\int_{t_0}^{t}e^2(t)dt, \tag{5}$$

where

$$e(t) = y_p(t) - y_m(t) \tag{6}$$

and

$$\frac{Y_p(s)}{U(s)} = K\frac{N(s)}{D(s)}e^{-\tau_p s}, \tag{7}$$

$$\frac{Y_m(s)}{U(s)} = K\frac{N(s)}{D(s)}e^{-\tau s}. \tag{8}$$

Using Massachusetts Institute of Technology (MIT) adaptive law, we obtain:

$$gradJ = \frac{\partial J}{\partial \tau} = \int_{t_0}^{t}e(t)\frac{\partial e(t)}{\partial \tau}dt. \tag{9}$$

The scanning step is λ, and τ is designed based on the performance indicators in the negative direction gradient.

$$\tau = -\lambda.gradJ = -\lambda\int_{t_0}^{t}e(t)\frac{\partial e(t)}{\partial \tau}dt, \tag{10}$$

$$\dot{\tau} = -\lambda e(t)\frac{\partial e(t)}{\partial \tau}. \tag{11}$$

Using (9) and y_p, which is independent of τ, we obtain:

$$\dot{\tau} = \lambda e(t)\frac{\partial y_m(t)}{\partial \tau}. \tag{12}$$

From (10) and (11) we obtain:

$$\frac{Y_p(s)}{Y_m(s)} = e^{(\tau-\tau_p)S} \therefore Y_m(s) = Y_p(s)e^{-(\tau-\tau_p)S}, \tag{13}$$

$$\therefore \frac{\partial y_m(t)}{\partial \tau} = L^{-1}\left[\frac{\partial Y_m(s)}{\partial \tau}\right]. \tag{14}$$

From (12), (13) and (14) we obtain:

$$\dot{\tau} = -\lambda e(t)L^{-1}\left[sY_p(s)*e^{-(\tau-\tau_p)s}\right]$$

$$\therefore \dot{\tau} = -\lambda e(t)L^{-1}\left[SY_m(s)\right]\therefore\dot{\tau} = -\lambda e(t)\frac{dy_m(t)}{dt}$$

$$\therefore \tau = -\lambda\int_{t_0}^{t}e(t)\frac{dy_m(t)}{dt}dt + \tau_0. \tag{15}$$

Hu Lianhua, Li Xinping, Tang Wei

τ_0 is the initial value of the reference model, and generally taken as the maximum probable value of the object time delay. Equation (15) is the resulting adaptive law.

3 Simulation

The simulation model is shown in figure 3, without loss of generality, the controlled object is [8]:

$$G_p(s) = \frac{K}{Ts+1}e^{-\tau s}, \qquad (16)$$

where T and K are known and assumed constant throughout the process and τ is the pure delay time which is changing throughout the process. Let K=3, T=10 and use a maximum probable value of τ of 50, i.e. τ_0 =50.

From part 2 we obtain the structure of IMC as:

$$G_c(s) = (Ts+1)/K . \qquad (17)$$

To achieve a smooth system, we add a filter in front of the IMC, which is set as follows: [9-16]

$$G_f(s) = \frac{1}{4.5s+1} . \qquad (18)$$

The scanning step should be chosen to match the system, here we use λ =95. The lower part of the simulation model graph expresses the adaptive law implementation calculated from formula (15). The simulation results are as follows.
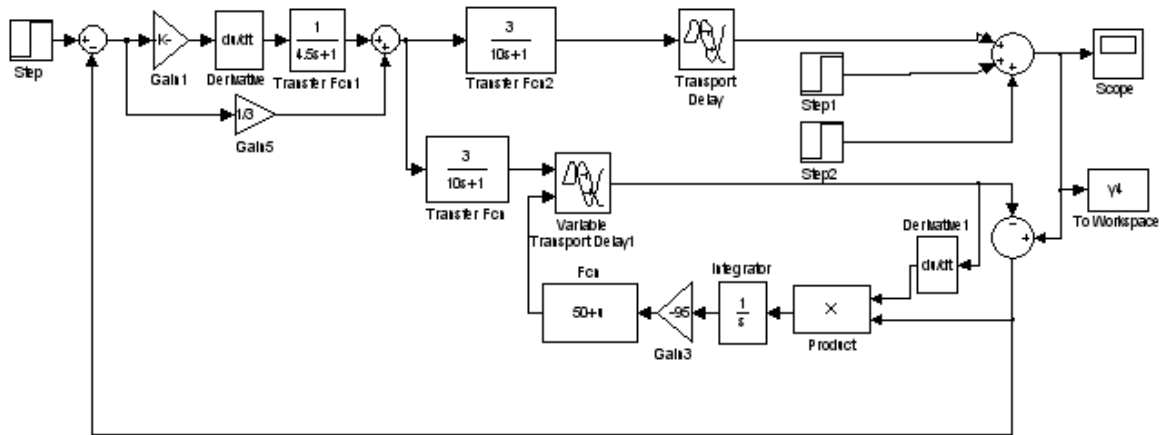
FIGURE 3 Adaptive IMC simulation model diagram
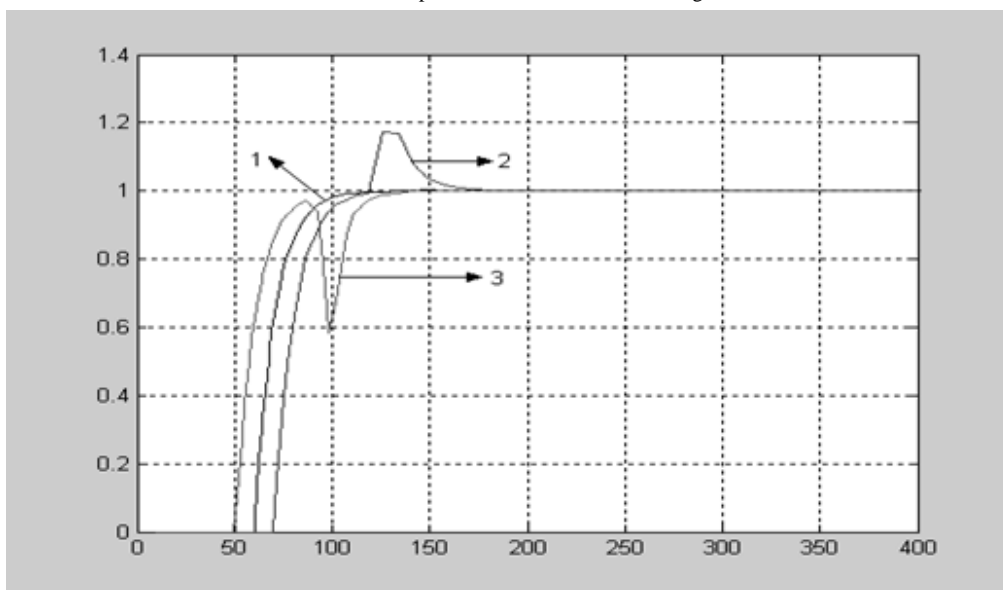
FIGURE 4 The simulation results of the adaptive IMC when the dead-time error is 20%
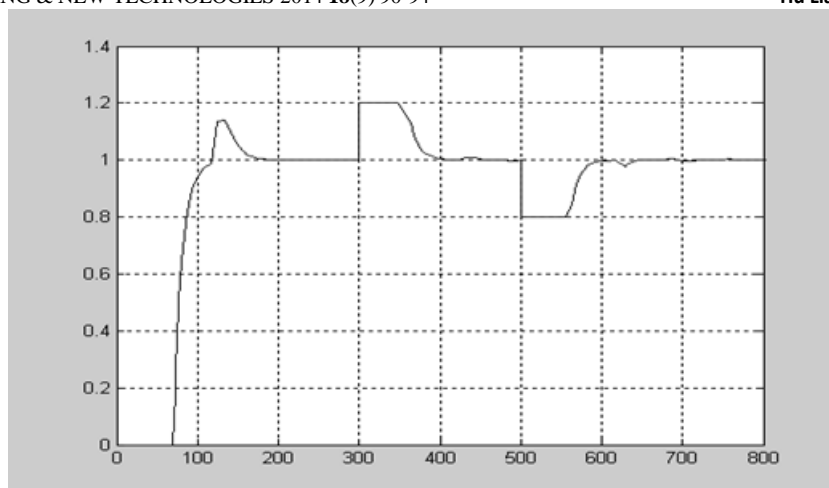
FIGURE 5 Adaptive IMC simulation results with dead-time and disturbance

Figure 4 displays the results of the designed adaptive IMC when the dead-time parameter is varying $\pm 20\%$. Curve 1 is the adaptive IMC model simulation result after parameter adjustment; it has the advantages of no overshoot and static error when the model matches the object. Even when the dead-time is reduced by 20%, curve 2 shows an overshoot of approximately 20%. When the dead-time is increased by 20%, the overshoot is 40% as shown in curve 3, but both conditions achieve stability rapidly without static error.

Figure 5 is the result of the adaptive IMC with both dead-time and disturbance. The dead-time is firstly increased by 20%. At 300s, the disturbance is increased by 20% and at 500s, the disturbance is decreased by 20%. The simulation curve shows that the system becomes rapidly stable and displays almost no static error after the delay time.

As can be seen from the simulation results, the adaptive IMC can achieve stability after oscillation when the model dead time error is $\pm 20\%$. The system can also achieve stability quickly under conditions of disturbance after the system's dead-time. Therefore, the designed adaptive IMC for variable parameters delay system has a suitable control effect.

4 Conclusions

In systems with large delays due to dead-time, the output signals in a conventional controller do not adjust fast enough to reflect the system's running conditions and therefore do not achieve a satisfactory control effect. The complexity and uncertainty of industrial processes causes frequent changes to object parameters; predictive control will not provide an adequate control effect in these situations, as it demands high model accuracy. This paper designs a model based on a first-order plus time delay adaptive law with local parameter optimization theory to identify the object parameters and modify the object online. The designed adaptive IMC model overcomes adverse effects that occur in the traditional IMC model due to inaccuracies, leading to improved controller performance, while maintaining the IMC characteristic of static error removal, ensuring that the system has good steady-state performance.

The simulation results show that the adaptive IMC has a low requirement for model accuracy, better robustness, good anti-interference ability and no steady-state error compared with conventional control methods. It is a superior control scheme for systems with variable parameters and a time delay.

References

[1] Jin Qibing, Liu Qie, Wang Qi, Tian Yuqi, Wang Yuanfei 2013 PID Controller Design Based on the Time Domain Information of Robust IMC Controller using Maximum Sensitivity *Chinese Journal of Chemical Engineering* **21**(5) 529-36

[2] Jin Qi B, Liu Q 2014 IMC-PID Design Based on Model Matching Approach and Closed-loop Shaping *ISA Transactions* **53**(2) 462-73

[3] Prakash Prashant, Verma Nishchal K, Behera Laxmidhar 2014 Eigenvalue Assignment via the Smith Predictor Based IMC-PID *IFAC Proceedings Volumes* **3**(PART 1) 997-1002

[4] Kansha Yasuki, Jia Li, Chiu Min-Sen 2010 Adaptive IMC controller design using linear multiple models *Journal of the Taiwan Institute of Chemical Engineers* **41**(4) 446-52

[5] Krause James M, Khargonekar Pramod P, Stein Gunter 1992 Robust adaptive control: Stability and asymptotic performance *IEEE Transactions on Automatic Control* **37**(3) 316-31

[6] Chien I-L, Fruehauf P S 1990 Consider IMC Tuning to Improve Controller Performance *Chemical Engineering Progress* **86**(10) 33-41

[7] Zhao Yao 2002 Internal Model Control Development Summary *Chinese Journal of Information and Control* **29**(6) 526-31

[8] Vilanova R 2008 IMC based Robust PID design: Tuning guidelines and automatic tuning *Journal of Process Control* **18**(1) 61-70

[9] Ali Ahmad, Majhi Somanath 2009 PI/PID controller design based on IMC and percentage overshoot specification to controller setpoint change *ISA Transactions* **48**(1) 10-15

[10] Tang Wei, Niu Xuzhong, Shan Wenjuan 2012 Application of IMC-PID in the Hydraulic Headbox Total Pressure Control *Advanced Materials Research* **462** 789-95

[11] Saxena Sahaj, Hote Yogesh V 2012 A simulation study on optimal IMC based PI/PID controller for mean arterial blood pressure *Biomedical Engineering Letters* **2**(4) 240-8

[12] Liu Tao, Gao Furong, Wang Youqing 2010 IMC-based iterative learning control for batch processes with uncertain time delay *Journal of Process Control* **20**(2) 173-80

[13] Bettayeb M, Mansouri R 2014 Fractional IMC-PID-filter controllers design for non integer order systems *Journal of Process Control* **24**(4) 261-71

[14] Maâmar B, Rachid M 2014 IMC-PID-fractional-order-filter controllers design for integer order systems *ISA Transactions*

[15] Petrovic T B, Ivezic D D, Debeljkovic D L J 2000 Robust IMC controllers for a solid-fuel boiler *Engineering Simulation* **17**(2) 211-24

[16] Kozub Derrick J, MacGregor John F 1989 Optimal IMC inverses: design and robustness considerations *Chemical Engineering Science* **44**(10) 2121-36

Authors

Lianhua Hu, born in 1978, Wei yang Area, Xi'an City, Shaanxi Province, P.R. China

Current position, grades: the lecturer of School of Shaanxi University of Science & Technology, China.
University studies: received his B.E in Electrical Engineering and Automation from Shaanxi University of Science & Technology in China. He received his M.E from Shaanxi University of Science & Technology University in China.
Scientific interest: His research interest fields include process control, control theory and control engineering
Publications: more than 10 papers published in various journals.
Experience: He has teaching experience of 13 years, has completed 3 scientific research projects.

Xinping Li, born in 1968, Wei yang Area, Xi'an City, Shaanxi Province, P.R. China

Current position, grades: the professor of School of Shaanxi University of Science & Technology, China.
University studies: received his B.E and M.E in Pulp and paper engineering from Shaanxi University of Science & Technology in China. He received his D.E from South China university of technology in China.
Scientific interest: His research interest fields include Cellulose functional materials, low pollution and low energy consumption of pulp and paper technology
Publications: more than 40 papers published in various journals.
Experience: He has teaching experience of 23 years, has completed 8 scientific research projects.

Wei Tang, born in 1971, Wei yang Area, Xi'an City, Shaanxi Province, P.R. China

Current position, grades: the professor of School of Shaanxi University of Science & Technology, China.
University studies: received his B.E and M.E in Electrical Engineering and Automation from Shaanxi University of Science & Technology in China. He received his D.E from Shanghai Jiao Tong University in China.
Scientific interest: His research interest fields include process control, control theory and control engineering
Publications: more than 50 papers published in various journals.
Experience: He has teaching experience of 20 years, has completed 10 scientific research projects.

Research on construction of normative ideological instruction teaching management system based on ZigBee wireless sensor network

Weili Pan

Zhejiang University City College, Hangzhou, Zhejiang, 310015, China

Abstract

In the process of standard ideological and political teaching management, teaching quality management, teacher management, student management and education environment management etc are all very important links. In order to establish and perfect the management system of ideological and political teaching in school, management utility should be developed. And the system optimizing should be comprehensively enhanced in teaching. We are to construct a management system on the basis of ZigBee through wireless sensor, RMON (remote monitoring) of ideological instruction teaching supervision, multimedia teaching and classroom environment management etc. will be designed through coordinating with Go-Ahead technology. The corresponding hardware and software facilities will be designed and developed. The role of wireless sensors in teaching management will be reflected through the image of flow image. Standardized teaching management would be genuinely realized with its advantages of convenience, utility, high efficiency, and energy conservation.

Keywords: teaching management, wireless sensor, ZigBee, RMON (remote monitoring)

1 Introduction

Teaching management includes plan management, teaching target management, the didactical process management, quality control, teacher management, student management ,teaching archives management .Teaching process is bilateral activity process composed of teacher's teaching and students' learning according to certain social requirement ,teaching purpose and the characteristics of students' physical and mental development. This process is composed of such elements as teachers, students, teaching content and means etc. The management of teaching process is also to determine the order of teaching work according to the rule of teaching process. The activities process of teaching objectives is realized through establishing corresponding method and such measures as plan, entertaining, check and summarization. However realizing them through relying on manpower merely with above management method is not enough obviously. With the rapid development of wireless sensor network in current world, and they have merged into various industries effectively. This paper is to suggest bringing the concept standardization of teaching management into wireless sensor system, thus better and more effective supervision and management will come true. This paper is to construct the system on the basis of ZigBee wireless sensor network [1], this is an intelligent system constructed by wireless communication technology ZigBee and flushbonading Web server Go-Ahead technology in short distance.

2 The introduction of ZigBee technology and go-ahead technology

2.1 ZIGBEE TECHNOLOGY

ZigBee technology is a kind of software protocol, which takes IEEE802.15.4 as the basis, and takes network security and application as guidance [2]. ZigBee is a kind of wireless communication technology with short distance, low transfer rate low energy consumption, low complexity and low power consumption [3]. It has high scalability and reliability. ZigBee is widely applied in various industries; especially it is prominent in consumer electronics, automotive automation and industrial control. It also has certain effect in medical equipment, home gateway and the enterprise gateway. Its development prospect is very broad. ZigBee protocol stack is very simple, its realization is relatively easy, and the needed system resources are less occupied. Complete ZigBee protocol stack is composed of physical layer, medium access control layer, network layer and application layer. The physical layer and MAC are defined by IEEE802.15.4, the network layer and application layer are defined by ZigBee alliance.

ZigBee takes independent work nodes as basis; it is a kind of wireless network type in stellate, cluster tree and mesh constructed through wireless network. The network is composed of three nodes: coordinator, router and terminal equipment. Coordinator is the centre network node, which is responsible of network composition and maintenance. Router is responsible of information

frame's route inside the internet; terminal equipment is to realize specific functional units [4]. ZigBee could also be divided into full function device and reduced function device, the nodes of reduced function device are only taken as the network terminal equipment. They are not communicated with each other, send and receive data through full function nodes .They do not have the function of routing and relay function. Full function device is used as three mid-nodes, which are responsible of communicating with all control child nodes, collecting data and releasing control, etc. ZigBee network has very strong reliability; this paper adopts ZigBee technology to improve instructional management system.

2.2 GO-AHEAD FLUSHBONADING WEB SERVER TECHNOLOGY

Flushbonading Web server is operated in embedded device on the basis of Web protocol server, it provides the condition and parameters information of embedded device in the form of webpage in computer, then transplants Web server into embedded system and accesses Internet. Flushbonading Web server could not only use browser to provide graphical user interface for users，realize the function of remote management and embedded monitoring system, but also make further development at server-side through combing with flushbonading Web service [5]. Flushbonading Web server has changed the remote control and management mode of embedded device, which does not needs, dedicated communication lines, transmittal information will not be limited to data information.

The common flushbonading Web server includes Https, Go-Ahead and Boa. The Iot gateway used in standardized education management system adopts Go-Ahead server, by which to realize the connection of ZigBee network and external Internet network，as well as information exchange and control. Go-Ahead is a kind of open source embedded Web server with more comprehensive function. It supports ASP, flushbonading JavaScript, CGI as well as HTML format of static page. The server could give fast response, process more than 50 requests for every second, and support a variety of operating systems such as Vx-works, Linux, WinCE and so on.

CGI (Common Gateway Interface)'s common gateway interface is a kind of communicating tool between Web server and browser. CGI procedure makes the webpage have interaction function; its procedure must operate on Web server. Most of the CGI procedure is used to explain and process the input information from browser forms, and corresponding processing would be generated in server or it will feedback corresponding information to browser.

3 The construction of standardized teaching management system

This paper is to realize the standardization of teaching management, construct a set of advanced administration system with high efficiency and low investment through using wireless sensor network model. As a kind of relatively intelligent supervisor mode used to replace traditional mode of labour management, it is able to create a new learning environment for students and teachers, as well as management system.

The system is composed of the students and teachers' handheld one-card, information acquisition node in classroom, wireless sensor coordinator, gateway server in classroom, web access terminal, and executing agency and so on. It could also be comprehended as typical B/S structural ZigBee system of perception layer, network layer, application layer [6].Wireless sensor nodes and the processing chip of coordinator adopt TI Company's CC253, and wireless communication adopts ZigBee technology. Intelligent classroom gateway takes Coretex －A8 processor and Linux operating system as the core, flushbonading Web server is constructed through adopting Go-Ahead.

The information nodes of wireless gathering in classroom include temperature, humidity, illumination and voice etc. Wireless execution node has light switch, air-conditioning switch, curtain switch, multimedia projector switch and so on，all these equipment could accept the order of control centre. According to the parameters such as intensity of light, the number of people in classroom etc., the brightness of the light will be of automatic adjustment, controller would make automatic adjustment. Controller could rate the classroom according to the noise level in the period of individual study; his handheld device designated with RFID tag should be available both in class and after class. The gateway of classroom accesses information about students and teachers makes statistical rating for the information of attendance, lateness and early leaving. Teachers could carry on the statistics for students' grade through the statistical functions of the system, thus the management will be convenient and real-time communication would be available, as it is shown in Figure 1.
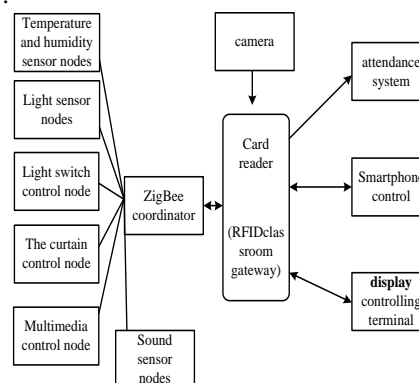
FIGURE 1 The structure of standardized teaching management system

In this system, ZigBee network topology model is of stellate type. ZigBee coordinator is the organizers of the network, which is responsible of network establishing and message routing. The sensing information in classroom could be of wireless access and sent to classroom gateway through serial port. User request could be received at Go-Ahead server-side and information data could be processed through CGI procedure, then the processed information will be of feedback to the user control and display terminal.

4 The Realization of hardware system and software system

4.1 THE HARDWARE OF TEACHING MANAGEMENT

Hardware facilities involved in this paper includes classroom gateway and ZigBee wireless sensor node .The gateway of classroom is the transfer station of interchanging all kinds of information data, attendance data, image data, audio data and remote display control terminal, statistical data of attendance system performance. All wireless node data, RFID data and camera data should be sent to gateway, which performs protocol conversion to data, and then it is transmitted to network display control terminal through Internet, or sent to handheld device through GPRS. The control data transmitted through Internet or GPRS by display control terminal will be equally sent to all wireless control nodes after address and protocol conversion. The gateway hardware adopts microprocessor Samsung S5PV210, this chip is based on ARM-CortexTM-A8 core, dominant frequency 1GHz.Peripheral resource includes modules of 512MB internal storage,1GB NAND Flash, 7 inch LCD touch screen resistance, ZigBee transceiver, RFID card reader, camera.

The communication module of ZigBee applied by us in teaching management is CC2530.It is a solution of real system on chip (SoC) applied for 2.4GHz IEEE802.15. 4, ZigBee and RF4CE [7]. In the design of overall ZigBee network, how to reduce power dissipation, especially reducing the power consumption of terminal nodes is the key of consideration. A typical terminal node of wireless sensor network is usually composed of sensor module, processor module, wireless communication module and power module. Processor module and wireless communication module adopt CC2430 chip, which would greatly simplify the design of radio-frequency circuit. As it is shown in Figure 2 and 3, the information acquisition module of sensor takes STM8S103F3 as microprocessor, the sensing information will be sent to CC2530 through AD transition and serial port.
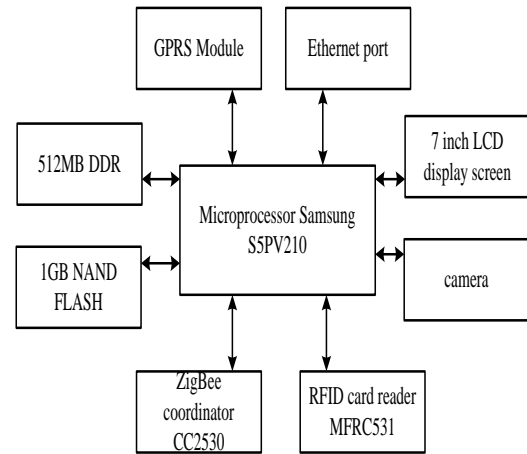
FIGURE 2 The structure of gateway hardware
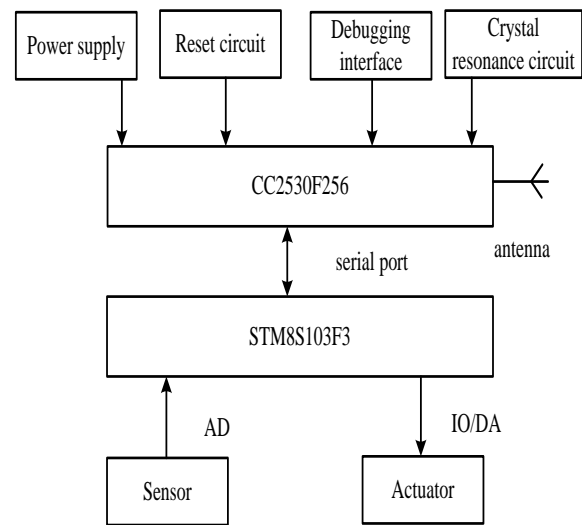
FIGURE 3 The structure of node hardware

From Figure 3, we can see that the effects of RBF neural network improved by two algorithms are better than that of normal PID control. In starting process of electromotor, it can effectively avoid starting overshoot, which is brought by normal PID control. Moreover, the time needed for the electromotor from start to stable work becomes shorter, thus improve the electromotor's working efficiency.

4.2 TEACHING MANAGEMENT SOFTWARE

The software facilities involved in this paper include classroom gateway, ZigBee coordinator and ZigBee node program design.

4.2.1 The design of classroom gateway software

The software systems' development of gateway is as following process:

(1) The transplanting of bootstrap program. Boot-loader is the first section of the program operated on hardware system after power-on. It is to realize the initialization of hardware equipment and establish the mapping of memory space, prepare good environmental

for calling operating system nucleus and user program. The Boot-loader used in this system is uboot.

(2) The transplanting of flushbonading operating system. Flushbonading operating system Linux has advantages of source code development, good stability, good transferability, powerful network function and so on, thus this system kernel is transplanted into the gateway of system. After downloading linux-2.6.35.7. tar.bz2, the kernel configuration file is modified, then cross compiling is performed and compressed, finally it is downloaded to the flash of gateway platform after generating uImage core files.

(3) The transplanting of root file system. The first loaded file system when starting operating system nucleus is root file system. Commands and tools in the file system are generated through applying busybox, and bin, sbin, user catalogue are also generated, then subdirectory of dev, etc, lib, proc, tmp, var, mnt, home are created. Files of inittab, rcS, fstab, profile etc. are prepared to be started under etc catalogue. Then all of the catalogue files are made into root file system file-yaffs format through using mkyaffs2image tool, finally it is downloaded to flash.

(4) The transplanting of flushbonading Web server – Go-Ahead. Its structure of catalogue could be observed through downloading Go-Ahead source packages and decompression. Transplant subdirectories containing a variety of operating systems include CE, ECOS, LINUX, LYNX, MACOSX, NW, QNX4, VXWORKS, WIN. Web catalogue could store html file, executable file of CGI cgi-bin procedure is stored under catalogue. In each operating system directory of Go-Ahead source code, there is main.c file, which is the entrance of overall system procedure and responsible of accomplishing server initialization, service monitoring and setting of relevant interface. Figure 4 represents the work process.
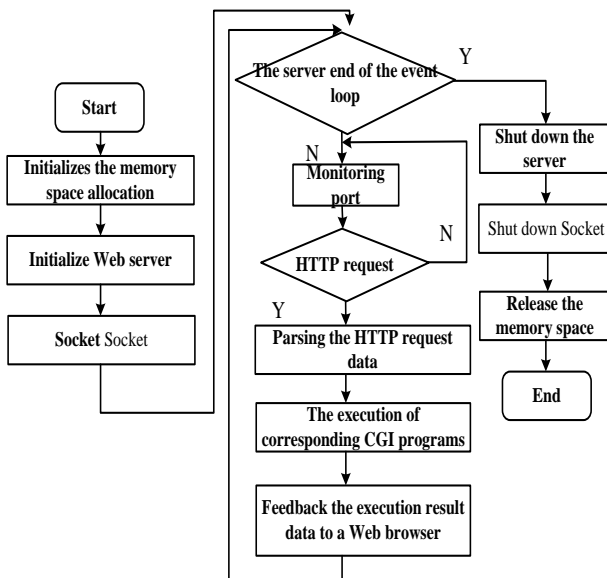
(5) The transplanting of gateway application program. The gateway of this system could not only act as Web server, but also take charge of directly indicating various information parameters of classroom and controlling various actuating equipment of classroom on gateway LCD at the same time. The GUI control program in this system could be accomplished by Qt.

4.2.2 The design of ZigBee coordinator software

The task of ZigBee coordinator linked with gateway serial port of classroom includes initializing CC2530F256 and protocol stack, constituting ZigBee network with the information sensing nodes of distributed temperature and humidity, illumination intensity, sound transducer etc.; detecting ZigBee wireless signal. If the network has route or there is terminal node entering network, then the address of network will be allocated to nodes. Network terminal node data will be accepted and transmitted to classroom gateway through UART, the control information sent by gateway will be accepted. Specific ZigBee terminal control node will be constructed after recombining data analysis [8]. It is shown as Figure 5.
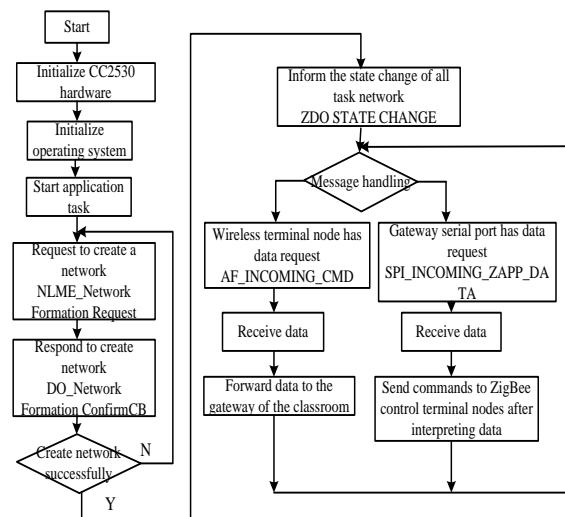
FIGURE 5 The workflow of coordinator software

4.2.3 Zigbee terminal node and remote control interface

The main work of ZigBee terminal node in this teaching management system is to apply for entering ZigBee network and performing data communications with ZigBee coordinator. Software workflow of terminal node is shown as Figure 6.

In the remote display control interface, any display control terminal could access the address of gateway through browser. Administrative staff with related jurisdiction not only could check parameters' information of specified classroom through verification of user name and password, but also realize the control of various equipment's in classroom. It is shown as Figure 7.

FIGURE 4 GoAhead workflow chart
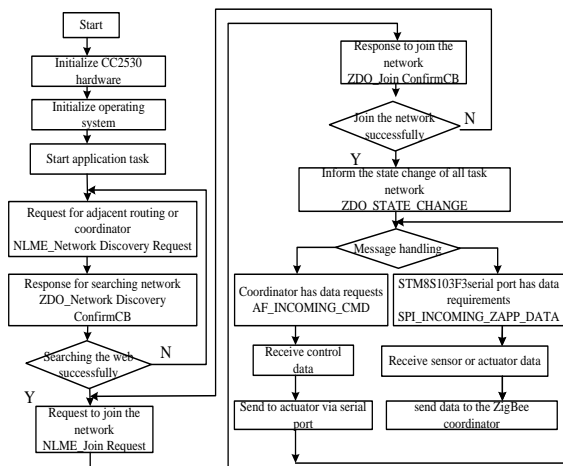
FIGURE 6 Workflow of terminal node software

FIGURE 7 Control interface of remote classroom

6 Conclusion

We are to combine wireless ZigBee technology and flushbonading Web server –Go-Ahead with low cost and low energy consumption, and then apply them in the system of standard teaching management. Through performing experiment on this equipment in ideological and political class, the writer has not only realized the remote management of classroom equipment, teachers' and students' attendance, discipline of classroom etc. through Internet, but also substitute traditional wired network with wireless sensor in obtaining classroom information. Thus, system application flexibility is enhanced; the efficiency of teaching management is strengthened. The experiment results indicate that technology based on ZigBee has represented certain feasibility in the management of standard education teaching management. Its performance is good and reliable, energy saving, convenient，it is a set of wireless control system with high quality and worthy of popularization.

References

[1] Yi Zheng 2010 Construct wireless sensor network based on ZigBee technology *Journal of Xiangfan University* **8**(31) 35-7
[2] Wenzhong Li, Zhaoyu Duan et al 2007 *Introduction and actual combat of ZigBee wireless network technology* Beijing: Beijing University of aeronautics and astronautics press
[3] Xue Han 2011 The research for the systematic design of smart campus based on ZigBee technology *Engineering of electronic design* **19**(24) 41-3
[4] Ling Jia, Xinyu Wang, Shujun Zheng et al 2011 *Principles and practices of ZigBee and wireless sensor network* Beijing: Beijing University of aeronautics and astronautics press 102 - 3

[5] Decai Wang, Du Xu, Yongping Jiang et al 2012 The research for intelligent housing system based on ZigBee and flushbonading server *Information technology* (9) 36 -9
[6] Qin Li, Lin Du et al 2013 The research and implementation of campus network environment's monitoring system based on ZigBee *Computer Knowledge and Technology* **29**(9) 6511-4
[7] Liting Cai, Ping Chen 2012 The design for ZigBee data acquisition system based on CC2530 *Computer Technology and Development* **22**(11) 197 － 200
[8] Yuexia Han, Wenfei Yang, Sumin Yang et al 2012 Flushbonading home intelligent control system based on ZigBee and GPRS *Information technology* (12) 102 - 8

Authors

Weili Pan, born in 1980, Zhejiang Province of China

Current position, grades: lecturer
University studies: Master's degree was earned in major of management, Tianjin University in 2006.
Scientific interest: educational administration

Hotel room design based on face recognition, environment monitoring and AV regulating system

Ning Liu*, Cheng Bin

Qingdao Hotel Management College, Qingdao, Shandong, 266100, China

Abstract

With the improvement of people's living standard, the tourism industry has transformed from the traditional sightseeing tourism to the vacation tourism, In order to improve the quality of the overall experience of resort hotel, the article by using the image acquisition technology, the facial expression recognition technology, the changes of Gabor and the DIB support, VC++ of bitmap provided by Visual Studio, has designed the basic member variables required by bitmap processing packaging DIB and the unrelated CDib of the member function device, and has constructed the facial expression recognition module of the mood regulation system of hotel interior. On the side of audio stream, use PC port as the audio stream scheme of the sound source, in the PC port, separate Av signal, preset the operation and treatment on the end of the playback software, and make self-regulation of sound adjustment in the output end of sound. The highly fidelity of sound has been established.

Keywords: hotel interior, Influence acquisition, feature extraction, frequency

1 Introduction

With the improvement of people's living standard, people's selection of tourism types has shifted from the sightseeing tourism to the vacation tourism. What it has brought about afterwards, is the booming of resort hotels. As people's entertainment venues, the resort hotel needs richer and more colourful and vibrant space than ordinary hotels, as it needs to meet people's growing demand for vacation and to create the life of idyllic beauty for people. People go for holiday from the busy and heavy work, it requires resort hotel also have the ability to adjust the mood, which provides a good opportunity for the designers to give play to the imagination and creativity [1].

The present domestic study shows that the environmental psychology requires the indoor environment design meet people's behaviour mode and psychological characteristics, as well as the personality of users and mutual adjustment of environment [2]. At this point, the paper has made a design proposal for the resort hotel interior designing by regulating the indoor environment to adjust the mood. It saves a lot of renovation costs for the hotel also while allows users to experience a different style of decoration.

2 Technical route for system implementation

Of which, (1) output captured image; (2) comprehensive analysis on weighted environmental mood parameters and weighted facial mood parameters; (3) analysis and calculation for the schemes of living environment suitable to improve mood; (4) process superior orders and invoke the video database, output configuration scheme; (5) implementation of program, projection image and play background music.
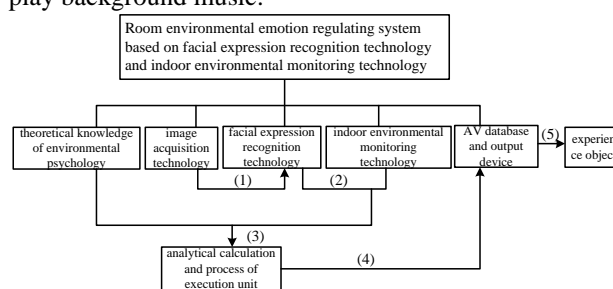
FIGURE 1 System technology route map

The system obtains the facial image of the participators with the scope that the images are able to be captured, with the image acquisition technology and input it into the image processing unit of the next level to conduct facial image analysis, capture the expression details with the facial expression recognition technology, analyse these feature values and give intelligence score Score-1 for the mood state presented by these expressions. Score-1 plus is the weighted facial mood parameters, which comes into being by Score-1 intelligence grades [3]. The system will analyse the image acquisition of the facial expression, while conduct acquisition and analysis of the indoor environment parameters. After the sensor has perceived the environment parameters, it will output these parameters to the weighted arithmetic of the next level. During the computing, these parameters will follow the predetermined algorithm to process the values of the light

---

* *Corresponding author* e-mail: 42895381@99.com

intensity, temperature, humidity and atmospheric pressure, to figure out Score-2 plus, namely the weighted environmental mood parameters. After the weighted facial mood parameters Score-1 plus and weighted environment mood parameter Score-2 plus, the system will conduct comprehensive analysis and computation and figure out the living environment schemes suitable to improve the mood, based on relevant rules of environmental psychology.

3 Key of system technology

3.1 FACIAL EXPRESSION RECOGNITION PROGRAMMING MODULE

The expression features extraction is realized by the use of Gabor transform. The Gabor transform belong to the windowing Fourier transform, which is a special circumstance in the short-time Fourier transform when the window function is evaluated as the Gaussian function. Practically, it is still to request convolution of the two-dimensional image [5] and it can extract the related features in different scales and directions [6]. Gabor function is similar to the biological role of human eye, and is often used for texture recognition, which is the key that it can extract facial expression feature [7]. The facial expression classification algorithm can be realized the use of DIB related to bitmap, which is provided by Visual Studio. The digital image can be realized by the custom class CDib provided by VC++, which is contained in Visual Studio. The design of the device-irrelevant class CDib can package the basic member variables and functions needed by DIB bitmap processing [8].

It can use the VS2008 development environment, MFC to achieve interface. After the initialization of template, run the software and open the diagram to be processed, then the fine recognition of expression can be conducted. Now the accurate expression recognition of happiness, disgust, anger, fear, sadness, neutral and shock in the database has been completed. The system block diagram is as shown in Figure 2.
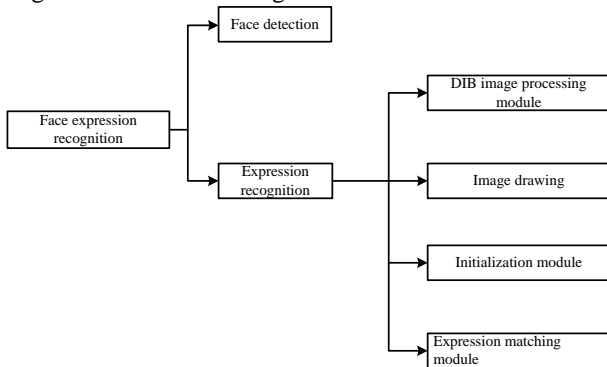
FIGURE 2 System block diagram

(1) Present block diagram of module debugged

The debugged DIB processing module includes the following aspects: save the DIB image, normalization

processing of image; constructor and destructor; construct DIB image; obtain CIB image from file; Obtain the subimage class and the both large-sized and small-sized DIB information; By taking pDIB as module, the model pGabor of Gabor as content, create new DIB file, of which both the length and the width is 1/5 of the original image.

(2) Initialization of module

The initialization module comprises constructor, the initialization of function and the release of resources.

3.2 AUDIO STREAMING SOLUTION

The audio stream is the mood and the weather data input by the previous system, and is also the final implementation scheme and the important component of output terminal of data stream in the hardware treatment, so it should be specially produced according to the needs and characteristics of the system. In consideration of the influence on the space shell material and sound system external structure and the blend use of data exchange interfaces of the central processing platform of PC end, decide to carry out two sets of predetermined audio stream solutions. In the stage of the cooperative work of various components in the system, and under the frame of the centre algorithm, integrate the demand of interface and resource and select the preferential audio stream solutions, which will be attached on importance later. Introduce in detail the two sets of predetermined audio streaming solutions:
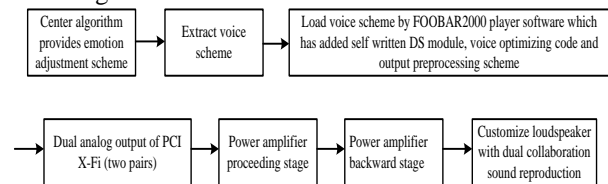
FIGURE 3 Schemes 1
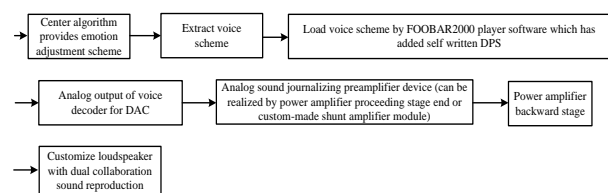
FIGURE 4 Schemes 2

The overall philosophy of the article of using PC end as the sound source is to process the separation of AV signal in the PC end, preset the operation and treatment of the output information on the top of the playback software, and make the self-regulated sound adjustment in the end of sound output. In this way, the audio stream becomes innovated and independent. Unlike the other main stream medias, the internal clocking of PC can synchronize and asynchronize the audio frequency and video frequency, PC can separate the audio stream with video stream and make digital decoding and simulated amplification of the audio stream, which is great emancipation for the audio effects in the audio end, and make the quality of sound effect more excellent. In the

level of audio decoder, the DAC module will be positioned that the external sound card or independent digital analogue converter will replace the inner on board sound card, built-in sound card, merged amplifier, etc. The advantage of it is that there will be inner disturbance and shielding problems, brought about by the built-in card and the on-board card and it solves the puzzle that it is hard to input and output multiple, which is caused by the previous and later levels with independent of merging efficacy. Of course the audio efficacy(device to merge the amplification video and audio streams), originally audio amplification device also has the typical functions of sound effect and fulu, which will conduct self-code-edition on FOOBAR2000 in the PC end, so as to preset sound effect, DSP setting and the operation of optimizing sound effect, so as to better achieve index parameters of sound channel isolation, delayed time range, all kinds of sound field modes, while it can separate the excessive operations and functions of the related part of videos in AV, which will affect the sound quality.

Different from the traditional independent type design of amplifier of PC together with AV power, the scheme will bring about the improvement in high fidelity. It requires the sound equipment to conduct distortionless amplification process of various indicators, during playback [9]. At the same time, use subtle music style for the dyeing of sound.

(1) When AV amplifier plays the sound source of large signal, the dynamic range is poor and the transient is lagged and it lacks of confidence; the audio stream makes the independent decoding he amplified and the process of dynamic state and frequency response more leisurely.

(2) With respect to AV amplifier routing, the amplifier module is the input and output of the same type interface and it does not need routing and fly line. The advantage of routing is reflected in shielding.

(3) The multiply interfaces displayed by traditional multimedia will affect the sound quality.

4 Conclusion

Based on the environmental psychology, with the environmental monitoring technology and the facial expression recognition technology, the article has conducted the acquisition of the related parameters and information on the hotel inner objects, made intelligent process and analysis on the relevant knowledge, worked out the adjustment scheme of hotel room environment suitable to improve the mood of the objects, and finally fulfil the purpose to improve the mood the target objects. From the detection and screening of weather factors by the integrated transducer and the reading of the facial image information by the facial expression recognition, the system will conduct intelligent analysis of the target objects, and then the actuating station piece will make changes to the environment hues, the wall patterns and background music. The system, by changing the tourist hotel indoor "scenery" to arouse the "affection" experiencing joy, and to form different mood themes, gives deep-felt care to the tourists with the recreational purpose and the need to get relaxed.

References

[1] Sun Tao 2005 *Research of design of public space to be built by resort hotel* Hunan, Hunan University
[2] Qiao Feng, Sun Yan 2007 Research and application of environmental psychology in the design of interior space *Architecture knowledge* **27**(3) 16-8
[3] Xiong Shuo, Wang Youli, Zheng Huang, et al. 2012 Design of mood regulation system of room environment *Electronic world* **2012**(19) 125-7
[4] Ye Gang 2007 Based on DS18B20 temperature control system design *Foreign electronic measurement technology* **26**(4) 31-3
[5] Hu Guosheng, Chen Yitian 2002 Fourier transform and wavelet transform general for - chirplet transform and Doppler transforms *Computer engineering and applications* **38**(5) 46-8

[6] Hu Guobiao 2011 *Research of remote sensing image scale effect based on registration analysis of* Chengdu University of Technology
[7] Ye Jingfu, Zhan Yongzhao 2005 Facial expression feature extraction based on Gabor wavelet transforms *Computer Engineering* **31**(15) 172-4
[8] Tang Yanyang, Zhang Qian, Bao liwei 2013 Research and realization based on VC image processing. wireless Internet *Scientific technology based on digital image processing VC* **2013**(3) 109-10
[9] Zhao Hongwei, Yue Bangyan 2005 AV power amplifier Hi - Fi power amplifier differences discussion *Household electrical appliances* **2005**(3) 60-60

Authors

Liu Ning, born on February 17, 1982, Shandong Province of China

Current position, grades: lecturer
University studies: Post-Graduate degree was earned in Qingdao University.
Scientific interest: tourism management, hotel management

Chen Bin, born on August 22, 1984, Shandong Province of China

Current position, grades: lecturer
University studies: Post-Graduate degree was earned in Qingdao University.
Scientific interest: tourism management, hotel management

The study of eccentric bunghole self-positioning system based on computer vision technology

Wei Yuan[1]*, Jianqi Zhang[2]

[1]*XianDian Universityr, Shannxi Xi'an 710071, China*

[2]*Weinan Normal University Shannxi Weinan 714099*

Abstract

This paper on the basis of deeply understanding the domestic automatic filling equipment's, develops an eccentric bunghole self-positioning system based on computer vision technology, through full automatic camera calibration and improved Hoff conversion algorithm to obtain the coordinate information of circle centre of eccentric fillers, thus controls the operation of driving stepper motor, realizes the automatic positioning of sprue guns to eccentric bunghole. The experiment proves the operation accuracy of this system is high, the speed is fast, the algorithm is efficient, stable, and a high practical value.

Keywords: computer vision technology, eccentric bunghole, automatic positioning, camera calibration, Hoff transformation algorithm

1 Introduction

Automatic filling machinery plays a very important role in modern industry automatic production, which is widely used in production fields such as petroleum, chemical engineering, medical treatment, beverage and so on [1]. However, when the filler of containers is eccentric entrance, the application of full automatic filling technology suffered serious restrict, at present the filling machinery of most enterprises adapts semi-automatic method, namely manual positioning barrel entrance filling. This becomes great dangerous operation sequence to the fillings of poisonous liquids with strong infiltration capacity such as sodium cyanide, hydrofluoric acid and so on. [2]. This paper on the basis of deeply investigating the current situation at home and abroad developing the eccentric bunghole self-positioning system based on computer vision technology, successfully solved the difficult self-positional problems which random distributed in circumference direction of eccentric entrances of filling lines, which cam guarantee the personal security of toxic liquid filling workers, can improve the mechanization and automation level of filling works, and with a good practicability [3].

2 The Eccentric Bunghole Self-Positioning System Based on Computer Vision Technology

This paper studies the eccentric bunghole self-positioning system based on computer vision technology is showed as Figure 1. The detected barrels are sent to filling station through transmission sequences, after computer obtains oil drum arrival information through optoelectronic switch, the industrial CCD vidicon installed above the oil drum can collect image information and send the images to image capture card in time. Image capture card processes digital decoding, A/D conversion to input signals, and through PIC bus sending the digital image data to computer memories to save. Computer makes analysis to the collected oil drum image information to calculate the centre positional coordinate in the images of top surface and send the information to electric control system, through line interpolation method, with the filler entrance of electric control system aims at oil drum spile to finish the automatic positioning of fillers. Finally the system control direct current machine stretches the fillers into spile through pinion and rack drive to realize automatic filling. Figure 2 lists the working process block diagram of eccentric bunghole self-positioning system based on computer vision.
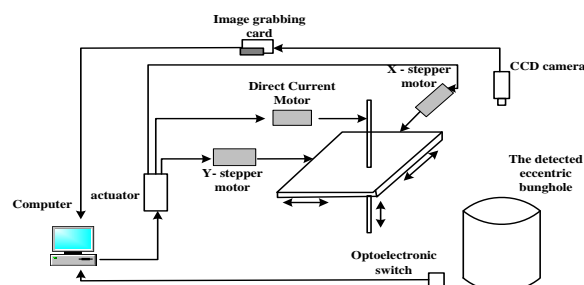
FIGURE 1 The simple diagram of eccentric bunghole self-positioning system based on computer vision

* *Corresponding author* e-mail:wnyuanwei@163.com

Yuan Wei, Zhang Jianqi

FIGURE 2 The working process block diagram of eccentric bunghole self-positioning system based on computer vision

2.1 CAMERA CALIBRATION [4] [5] [6]

The designed camera calibration of this paper does not need any artificial participation and all the calibration processes are completed automatically in order to adapt the automation-filling request of oil drum. The process includes six steps such as making threshold segmentation, parameter circular array mark, enter automatic withdrawal of parameter circular array and the vidicon parameter calculation based on least square fit to target images, which is showed as Figure 3.
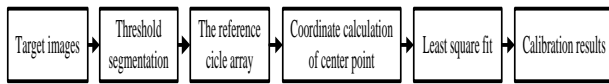
FIGURE 3 Vidicon calibration process diagram

2.2 COORDINATE SYSTEM IN CAMERA CALIBRATION

As it shows in Figure 4, the coordinate system in camera calibration has the following three layer's coordinate systems:

1. World coordinates of scene coordinates, the general three-dimensional scenes of (XYZ) are expressed by this coordinate system.
2. Camera coordinates. $\left(X_cOY_c\right)$ takes the focal length centre of microspore vidicon as the origin, takes the three-dimensional rectangular coordinate system of vidicon optical axis established on the basis of Z axle, namely the camera CCD plane coordinate system.
3. Pixel coordinates. $\left(X_fO_fY_f\right)$ also called frame saving coordinate system of computer image (fix on the images), which is the rectangular plane coordinate system in pixels, its origins locates on the top left corner of images.
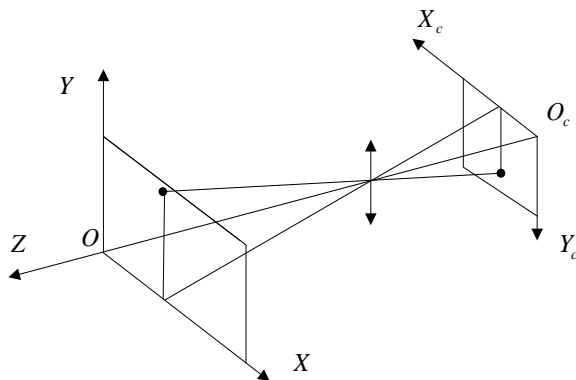
(a) World coordinate and camera coordinate
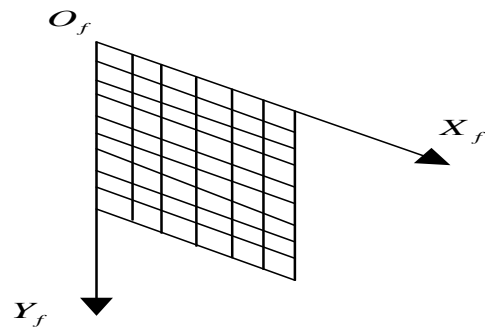
(b) Pixel coordinates

FIGURE 4 The three-layer coordinate system during vidicon imaging progress

(1)The Calibration Method of This Paper

Placing the showed target location of Figure 5 into the plane XOY of world coordinate system which showed as Figure 4(a),suppose the centre point coordinate of any reference circle is (X,Y), the imaging coordinate of this point within camera CCD plane coordinate system is $\left(X_c,Y_c\right)$. In image pixel coordinate supposed as $\left(X_f,Y_f\right)$, any point P(X,Y) in target with the coordinate in projection point $P_c\left(X_c,Y_c\right)$ should meet:

$$\begin{cases} X_c = KX \\ Y_c = KY \end{cases}. \tag{1}$$

K is proportional action factor in the algorithm.
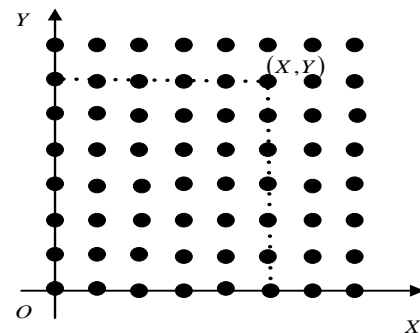
FIGURE 5 Calibration target
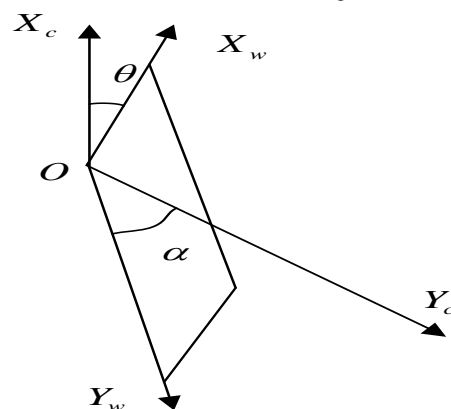
FIGURE 6 Rotation and decentration of camera coordinate system

Considering the factors such as camera optical axis out of plumb and CCD installation inaccuracy will lead to the rotation error and decentration error in CCD plane, which need to make conversion and correction to CCD plane. Suppose the CCD plane coordinate system after conversion and correction is $X_w O Y_w$ (see also to Figure 6), and then the rotation, decentration error of CCD plane can use the following formulas to correct:

$$\begin{bmatrix} X_w \\ Y_w \end{bmatrix}^T = \begin{bmatrix} X_c \\ Y_c \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \quad (2).$$

λ_x, λ_y in the formula means the horizontal equivalent and vertical equivalent, which respectively on behalf of the actual value of unit pixel in image plane $(X_f O_f Y_f)$ countdown to world coordinate system XYZ on horizontal equivalent and vertical equivalent?

Simultaneous formula (1), (2) to obtain:

$$\begin{cases} X_f = A_0 + A_1 X + A_2 Y \\ Y_f = B_0 + B_1 X + B_2 Y \end{cases}, \quad (3)$$

A_0, B_0 in the formula is the relative transform constant on horizontal equivalent and vertical equivalent.

$$\begin{cases} A_1 = \dfrac{\cos\alpha}{\lambda_x} \\ A_2 = -\dfrac{\sin(\alpha+\theta)}{\lambda_x} \\ B_1 = \dfrac{\sin\alpha}{\lambda_y} \\ B_2 = \dfrac{\cos(\alpha+\theta)}{\lambda_y} \end{cases} \quad (4)$$

In the formula, if A_1, A_2, B_1, B_2 is known, then we can get:

$$\sin\alpha = \sqrt{\dfrac{A_1^2 B_1^2 - A_2^2 B_1^2}{A_1^2 B_2^2 - A_2^2 B_1^2}}, \quad \lambda_x = \left| \dfrac{\cos\alpha}{A_1} \right|,$$

$$\lambda_y = \left| \dfrac{\sin\alpha}{B_1} \right|. \quad (5)$$

Suppose the coordinate measuring value of any centre of reference circle in target is $(X_i, Y_i), i = 0, 1, 2, ..., n$, the corresponding coordinate in the image coordinate system is $(X_{fi}, y_{fi}), i = 0, 1, 2, ..., n$, with the least square fitting method we can obtain the coefficient $A_0, A_1, A_2, B_0, B_1, B_2$ in formula (3).

Suppose $P = \begin{bmatrix} A_0 & A_1 & A_2 & B_0 & B_1 & B_2 \end{bmatrix}^T$, (6)

$$Q_i = \begin{bmatrix} 1 & X_i & Y_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_i & Y_i \end{bmatrix}, \quad (7)$$

$$Q = \begin{bmatrix} Q_1 Q_2 Q_3 ... Q_n \end{bmatrix}^T, \quad (8)$$

$$R_i = \begin{bmatrix} X_{fi} Y_{fi} \end{bmatrix}^T, \quad (9)$$

$$R = \begin{bmatrix} R_1 R_2 R_3 ... R_n \end{bmatrix}^T. \quad (10)$$

The n in the formula is the centre number of target reference, which participates in the least square calculation, according to formula (3):

$$QP = R. \quad (11)$$

Substituting (X_i, Y_i), (X_{fi}, Y_{fi}) sequence into (11) formula, with the least square calculation we can obtain:

$$P = (Q^T Q)^{-1} Q^T R. \quad (12)$$

From the above analysis we can know the parameter fitting method based on the least square can proceed as following steps:

1. Measuring the coordinate (X_i, Y_i) of each centre of target reference circle array in world coordinate system.
2. Adapting image processing method to measure the corresponding coordinate of (X_i, Y_i) in image coordinate system is (X_{fi}, Y_{fi}),
3. According to 1., 2. results to build matrix R and Q;
4. Getting column vector P, thus determines $A_0, A_1, A_2, B_0, B_1, B_2$;
5. After obtaining $A_0, A_1, A_2, B_0, B_1, B_2$, then according to the image information (X_f, Y_f) of measured objects within algorithm (2-86) to recognize the target object information (X, Y) in environment.

2.3 IMAGE CAPTURE AND PROCESSING [7]

When capture cards are at work, sends the collected data to memory. Through Read Form Mem() function can display the image data which saved in memory, through the time testing of system shows that when image capture cards are at work, collection to memory is about 500ms, the collection directly to screen will not more than 40ms. In order to accelerate collection speed, system sends the images which collected by image capture cards to VGA display card and displayed through screen, then

invoking ReadDisp Window() function to read the image data and save the image from screen, and then store to specified memory buffers, the image processing progress is showed as Figure 7:
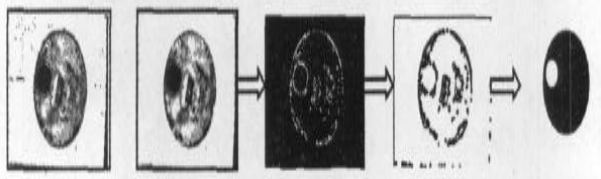
FIGURE 7 Figure collection and processing

2.4 DETERMINES THE CENTER COORDINATE POSITION OF ECCENTRIC BUNGHOLE WITH HOFF TRANSFORM ALGORITHM

The calculation of Hoff transform is very complex, which includes great number of squares, evolution operations. With the value range of r constantly enlarged in circular testing application, the size of three-dimensional group in parameter space increases proportionally, which needs to consume lots of computer memories and the efficiency becomes low. Therefore, this paper summarizes the subjects and proposes a new circular Hoff transform algorithm, and the basic steps are showed as follows:

1. Getting the binary edge image through canny edge testing;
2. Defining accumulator space of transform position, its size is in line with the size of binary edge image;
3. Looking for next edge point, with this pixel as the central confirming （2 k+ 1)×(2k+1) windows, and establishes coordinate system xoy in this edge point;
4. Counting the number of edge points in windows, if the number of point equals to (2k +1) goes to 5., otherwise to 3.;
5. Looking for endpoint A, B;
6. Verifying whether A, B, O these three points are collineation, if collineation goes to 3., otherwise calculates the perpendicular bisector function of segment AB;
7. Using the mutual positional relationship between origin O and segment AB to determine circular bending direction, namely the accumulation direction of centre in parameter position, and accumulates to this direction in parameter position.
8. Inspecting whether finishing the scanning of overall image, if not finish, then goes to 3., if finish goes to 9.;
9. When all the edge point detection finished, finding the maximum of accumulator results in parameter position, the corresponding accumulator coordinate of this value is the central coordinate of detected circles.

This Hoff transform algorithm can wipe off straight edge and noise point, refuse the meaningless accumulated information through using circular edge outline information, compared with traditional Hoff transform algorithm, its computational effort is small, the operation speed is fast, and occupied computer memory is low.

When uses this algorithm to determine bunghole central position, mainly needs to go through image acquisition, image cutting, edge detection and Hoff transform these four steps, which is showed as Figure 8. After excluding cover centre of oil drum, then can according to image coordinate value $\left(X_f, Y_f\right)$ and camera calibration coefficient we can obtain the world coordinate value (X, Y) in the central of eccentric bunghole.

(a) Original oil drum image

(b) Oil drum image cutting
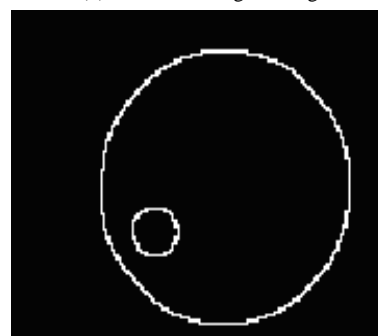
(c) Canny edge detection
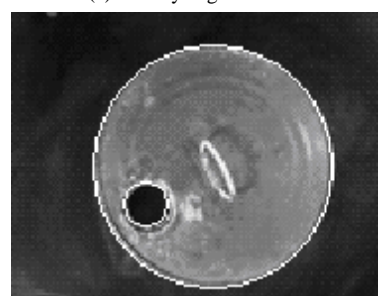
(d) Fast Hoff transform results

FIGURE 8 The central extraction process of eccentric bunghole

2.5 MOTOR DRIVE SYSTEM OF MARCHING TYPE DESIGN

After vision system detecting the central coordination information of eccentric bunghole, we need to send the bunghole central coordination to the above of fillers with computer and realize automatic self- positioning fillings.

This paper designs the marching type drive system with the core of L297.L298, the circuitous philosophy is showed as Figure 9.
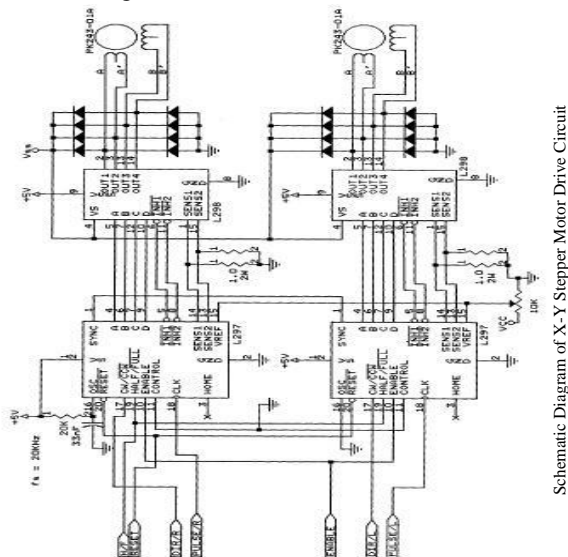
FIGURE 9 Circuitous philosophy figure of X-Y marching type drive

The core of stepping motor controller chip circuit L297 is pulse distributor, adapts plastic packaging form of 20 feet dual in-line type, usually with +5V to supply power. There are two PWM choppers in L297 for controlling winding current to realize constant current cut wave, thus obtained a better torque-frequency features, which applies to the control of bipolar two phases stepping motor or unipolarity four-phase stepping motor.L297 only need to receiver direction (forward / rollback), model(half step/basic step pitch), clock (step-by-step impulse) these three input signals from upper computer, which produces three kinds of phase sequence signals corresponding to three different working manners: half step method (four phase and eight step); single-phase incentive method (single phase and four step); double-phase incentive method (two phase and two step). The initial state of L297 is ABCD=0101.

L298 is double-H bridge high voltage large current power integrated circuit, it receives standard logic electrical level signals, which can use for the electric loading such as driving relay, coil, continuous current motor and stepping motor and so on.

The system consists of L297 and L298 drives two-phase stepping motors, the highest voltage is 46V, each phase current reaches to 2A.

The parallel end of computer sends clock signal, positive and negative turning signal, working pattern signal, enable signal and control signal. The steering of

CW/CCW control motor is opposite when evaluates 1 or 0.CLOCK is stepping impulse signal input end, on falling edge of each impulse, electrical machine generates a stepping. HALF/FULL is half step or basic step pitch mode setting, 1 is half step mode, and 0 is basic step pitch. If HALF/FULL evaluates low level, when impulse distributors are in odd number state, then is two-phase incentive mode; when impulse distributors are in even number state, then is single-phase incentive mode. When CONTROL chopper control is 0, controls INH1 and INH2, when is 1, controls ABCD.ENABLE enables input, when is 0, INH1, INH2, A, B, C and D all is 0.RESET input, when is 0, pulse distributors return to initial state. In circuit, the two resistances in fifteen feet of L297 used for adjusting the reference voltage of chopper circuit, which is compared to the feedback potential size through pin 13.14 to confirm whether to process chopper control, thus achieved to control the current peak of machine winding and the aim to protect stepping motor.

2.6 OPERATION EFFECT MONITORING OF SYSTEM

For adjustment, eccentric bunghole automatic positioning system sets up manual functions, X axis, Y axis both can realize inching. Bunghole circular image coordinate and world coordinate can be displayed real-time. The chose steeping angles of electrical machines α is 0.91°, which adapts belt drive, the diameter of gear pitch is 12.59mm, and pulse equivalent of stepping motor is 0.1mm.

The positional accuracy of computer vision eccentric bunghole automatic positioning system is the sum of eccentric bunghole circle recognition error 0.06mm and filling gun moving error 0.1mm,its value is 0.16mm.

Every time detects an eccentric bunghole ,the runtime of positioning system is the sum of eccentric bunghole circle recognition time and the time of machine control stepping motor drive aligns to eccentric bunghole circular, which is about 6.5 seconds.

Operating effect shows this the positional accuracy of this computer vision eccentric bunghole automatic positioning system is superior, running speed is fast, and realized conveniently and speedily, which has a practical importance to improve bottling equipment industry.

3 Conclusion

In order to solve the accurate positioning problem of eccentric bunghole during stepping bunghole filling, this paper develops an eccentric bunghole self-positioning system based on computer vision technology. At first makes a brief expound to system working process and working principle, then aims to the working process of this automatic self-positioning system to make analysis and explanation one by one: the vidicon calibration adapted by this paper dispense with artificial participation, all the calibration progress are completed

107

automatically, including target image, threshold cutting, reference circle array mark, automatic withdrawal of reference circle array mark and the vidicon parameter calculation based on the least square fitting these six steps, finally proposes a new Hoff transform algorithm. This algorithm uses circle edge outline information features directly deletes straight edge and noise point, as well as refuses accumulate meaningless image information. Compared with traditional Hoff transform, its accumulation parameter domain is two-dimensional array, the computational effort is low, operation speed is high and can save computer memory usage effectively; then this paper designs stepping motor drive system with the core of L297, L298, after this system receiving the eccentric bunghole information coordinate sent by computer, drive syringe removes to the upward side of filler, thus realizes automatic positioning filling. Finally, obtains the self-positioning system of computer vision eccentric bunghole has the advantages of high positional accuracy and fast operational speed, which has a widely practical significance and application value.

References

[1] Yan Lamei 2001 Current analysis of our country's packaging machinery and the formation of international market *China Packaging Industry* **15**(5) 4-8

[2] Liu Shuguang, Liu Mingyuan 2000 Machine vision and its application *Machine manufacturing* **38**(7) 20-2

[3] Barnea D J, Silverman H E 2003 A class of algorithm for fast digital imagine registration *IEEE Trans, Computers* **21** 179 - 1861

[4] Qin Maolin, Ma Songde, Li Yi 2000 Camera calibration summary in computer vision *Automation Journal* **26**(1) 43-55

[5] Armstrong M, Zisserman A, Hartley R I 1996 Self-Calibration from image triples *In: Proceedings of the Fourth European. Conference on Computer Vision, Lecture Notes in Computer Science* **1064** 3-16

[6] Fan Peifen 1985 Applied linear algebra *China meteorological press* **1** 94-8

[7] Ching M S, Suen Y, Kazuhiko Y 1992 Historical Review of OCR Research and Development *Proceedings of IEEE* **80**(7) 1029-58

[8] Xu L, Oja E, Kultaned P 1998 A new curve detection method: Randomized Hough Transform (RHT), Pattern Recognition Letters **11**(5) 331-8

[9] Yuen H K, Illingwoth J, Kitter J 1995 Detecting partially occluded ellipses using the Hough transform *Image and Vision Computing* **7**(1) 31-7

[10] Zhang Hongmin, Zhang Yujian 2004 Circle coordinate fast extraction method based on generalized Hough conversion *Journal of Jianghan petroleum institute* **22**(2) 64-6

Authors

Wei Yuan, born in 1973, Shaanxi Province of China

Current position, grades: PhD Candidate, associate professor
University studies: Bachelor of science degree was earned in major of the application of electronic technology, Shaanxi Normal University in July, 1997; master of engineering's degree was earned in major of science and technology, Xidian University in April, 2007.
Scientific interest: infrared image system simulation and evaluation

Jianqi Zhang, born in 1960, Shaanxi Province of China

Current position, grades: professor, PhD candidate supervisor
University studies: Bachelor of engineering's degree was earned in major of physics, Xidian University in 1982; master's degree was earned in major of physical and electronic components in Xidian University in 1987; PHD degree was earned in major of microelectronics science and technology, Xi'an Jiaotong University in 1998.
Scientific interest: photoelectric imaging system simulation and evaluation

The analysis and avoidance of fault agent in flocking of multi-agent

Yutian Liu*, Junjie Hu

Faculty of Electronic and Information Engineering, Zhejiang Wanli University, Ningbo, 315100, China

Received 1 May 2014, www.cmnt.lv

Abstract

In the last decade considerable research efforts have been spent to the motion of flocking of multi-agent. Special attention has been put in the applications, especially for those operations in real environment where a high degree of safety as well as self-diagnostics capabilities are required. The development of effective strategies of fault diagnosis for flocking of multi-agent is a critical research task. In this paper, the flocking motion of multi-agent with a leader is studied. When flocking in the real environment, it is inevitable for agents to occur faults. The faults occurred in different agent will lead to different effects for flocking. According to the variety of the velocity of agents, the fault types are classified. A fault agent avoidance method is proposed and implemented in a multi-agent flocking system. The simulation results show the method can help the agents to avoid the fault agent.

Keywords: multi-agent, flocking, fault agent, fault agent avoidance, leader

1 Introduction

As advancing the research of flocking of multi-agent, the flocking motion is asked for higher requirements on its safety and reliability [1-5]. In reality, it is inevitable that some faults may happen to agents during the motion of flocking. It is the challenge that the research of flocking of multi-agent has to face. During the motion, multi-agent cannot form a stable formation and will continue unpredictably and erroneously, if the fault agent cannot be detected or other functional agents cannot avoid the fault agent in time. As a result, the flocking cannot complete the scheduled tasks successfully. Or more possibly, more serious damages may occur.

During the motion of flocking of multi-agent, once one of the agents has failure, the caused chain reaction will lead to the malfunctions happening to other agents. Therefore, the foundation of the application of flocking of multi-agent in real life is to analyze the effects of fault agent imposing on other agents, timely obtain a complete knowledge of the fault agent and correctly handle it, and guide other agents to continue the motion. But the existing literatures about flocking of multi-agent seldom focus on the problem of handling fault agents in flocking of multi-agent [6-9].

The fault in different agents of flocking will play distinct effects on the motion of flocking. Selecting the most effective approach to solve problems based on the natures of the fault agents is a good way for the flocking to accomplish the tasks.

This paper focuses on classifying and analysing the fault agents according to their roles and importance in the motion of flocking, and proposes a systemic approach for the multi-agent to avoid the fault agent.

2 Flocking of multi-agents with a leader

The topology structure of multi-agent system is represented by the neighbouring graph $G = (V, E)$ [9-11]. And the neighbouring graph is defined to be an undirected graph consisting of a set of vertices $V = \{n_1, n_2 L, n_n\}$, whose elements represent agents in the group, and a set of edges $E = \{(n_i, n_j) \in V \cdot V \mid n_i \sim n_j\}$ containing unordered pairs of vertices that represent neighbouring relations at time *t*. Considering N agents, the motion of each agent is denoted as,

$$\begin{cases} \dot{r}_i = v_i \\ \dot{v}_i = u_i \end{cases}, i = 1,2, ,N, \tag{1}$$

where, $r_i \in R^2$ is the position vector of Agent *i*. $v_i \in \sim R^2$ is the velocity vector of Agent *i*. $u_i \in \sim R^2$ is the control input acting on Agent *i*. $r_{ij} = r_i - r_j$ is the position difference vector.

In order to describe the spatial order of the desired configuration of flocking in a proper analytical framework, to avoid collision with other agents, each agent should keep a safety distance from its neighbours. Then, we define the coupling constraints as that for agent *i* and *j* [12, 13],

$$\left\| r_i - r_j \right\| = d, \tag{2}$$

where, $\left\| \cdot \right\|$ is the Euclidean norm. We assume that the relative distance between agents will be *d*. The speed and direction of movement of each agent will approach the same.

The fundamental idea of flocking with leader-followers model is that, in a group of multiple agents, one

***Corresponding author** e-mail: lyt808@163.com

is assigned as the leader and others are followers. The task of leader is to lead the entire formation to the destination. The task of follower is to reach the destination following the leader. The goal is that the follower tracks the leader with a desired distance and a desired relative speed. In this paper, one leader agent with 5 follower agents based formation is considered as indicated in Figure 1.
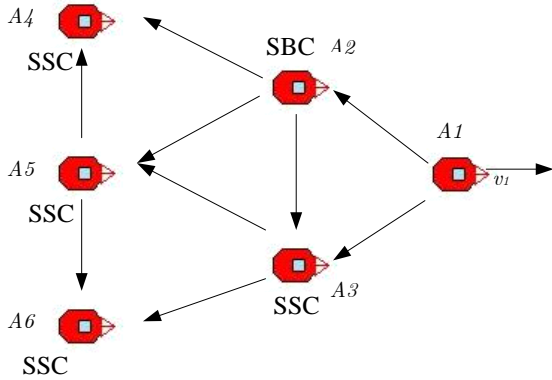
FIGURE 1 Flocking of multi-agent with a leader

Figure 1 shows the flocking model. *A1* (Agent 1) is the leader agent, and moves in a predefined trajectory. *A2* (Agent 2) and *A3* (Agent 3) are the follower agents. *A2* employs SBC strategy, so it only maintains its separation and bearing with respect to *A1, A3, A4, A5* and *A6* employ SSC strategy, maintaining its separation from *A1, A2* or *A3* [14-16]. The following and controlling relationship of the 6 agents is as following: $A2 \leftarrow A1$ (*A2* is controlled by *A1*), $A3 \leftarrow \begin{cases} A1 \\ A2 \end{cases}$ (*A3* is controlled by *A1* an *A2*),

$A4 \leftarrow \begin{cases} A2 \\ A5 \end{cases}$, $A5 \leftarrow \begin{cases} A2 \\ A3 \end{cases}$, $A6 \leftarrow \begin{cases} A3 \\ A5 \end{cases}$.

The dynamic equation of Agent 1 is given as follows,

$$\begin{cases} \dot{x}_1 = v_1 \cos\theta_1 \\ \dot{y}_1 = v_1 \sin\theta_1 \\ \dot{\theta}_1 = \omega_1 \end{cases}, \quad (3)$$

where, $[x_1, y_1, \theta_1]$ is the position of Agent 1, v_1 is the linear speed , ω_1 is the angular speed of Agent 1.

The dynamic equation of Agent 2 is given as follows,

$$\begin{cases} \dot{l}_{12} = v_2 \cos\gamma_1 - v_1 \cos\varphi_{12} + d\omega_2 \sin\gamma_1 \\ \dot{\varphi}_{12} = \dot{l}_{12}\{v_1 \sin\varphi_{12} - v_2 \sin\gamma_1 + d\omega_2 \cos\gamma_1 - l_{12}\omega_1\} \\ \dot{\theta}_2 = \omega_2 \end{cases}, \quad (4)$$

where, $\gamma_1 = \theta_1 - \varphi_{12} - \theta_2$, $[l_{12}, \varphi_{12}, \theta_2]$ is the position of Agent 2. v_2 is the linear speed, ω_2 is the angular speed of Agent 2.

The dynamic equation of Agent 3 is given as follows:

$$\begin{cases} \dot{l}_{13} = v_3 \cos\gamma_1 - v_1 \cos\varphi_{13} + d\omega_3 \sin\gamma_1 \\ \dot{l}_{23} = v_3 \cos\gamma_2 - v_2 \cos\varphi_{23} + d\omega_3 \sin\gamma_2 \\ \dot{\theta}_3 = \omega_3 \end{cases}, \quad (5)$$

where, $\gamma_2 = \theta_2 - \varphi_{23} - \theta_3$, $[l_{23}, \varphi_{23}, \theta_3]$ is the position of Agent 3. v_3 is the linear speed, ω_3 is the angular speed of Agent 3. The dynamic equations of other agents can be deduced by analogy.

3 Fault types in flocking of multi-agent

The velocity and position of each agent are vitally important in the motion of flocking. In this paper, faults are classified by the difference of the velocity of each agent. According to Figure 1, there are two types of agent, leader fault and follower fault. We classify the faults into two situations, leader fault and follower fault.

The definition of the two situations is explained as follows.

3.1 FAULT OF LEADER

In the motion of flocking, the task of leader is to lead the entire formation to the destination. Once the leader malfunctions, it will not be able to fulfil the task. The threshold value range of normal movement velocity of agent is set to be $[\varepsilon_1, \varepsilon_2]$. Then, three fault types are discussed according to the linear velocity of leader as follows.

1) $0 \le v_1^k < \varepsilon_1$, where v_1^k is the linear velocity of leader. $v_1^k = 0$ represents two possible cases. One case is the malfunction of the communication module of leader. Since other agents cannot receive the motion information of leader, they will consider the velocity of leader to be 0. Other case is the malfunction of the mechanical module, which will bring leader to a stop. In both cases, other agents will follow leader to a stop. This will cause the task to fail.

$0 < v_1^k < \varepsilon_1$ represents that the velocity of leader is below the normal range. Other agents will not be able to reduce the speed to follow leader, and leader will obstruct the motion of other agents. In this case, leader is considered to be faulty, and be the obstacle of other agents.

2) $\varepsilon_1 < v_1^k < \varepsilon_2$. In this case, although the velocity of leader is different from the normal one, it still in the threshold value range. In order to fulfil the task, other agents will adjust their motion speed to follow leader.

3) $v_1^k < \varepsilon_2$. In this case, the velocity of leader is beyond the threshold value range. Other agents cannot follow leader to move in so high speed. So, the leader is considered to be faulty. However, if the direction of motion is remained unchanged, the leader will not obstruct the motion of other agents.

3.2 FAULT OF FOLLOWER AGENT I (I=2,3,4,5,6)

1) $0 \leq v_1^k < \varepsilon_1$, where $v_1^k = 0$ represents two possible cases. One case is the malfunction of the communication module of Agent i. Since other agents cannot receive the motion information of Agent i, they will consider the velocity of Agent i to be 0. Other case is the malfunction of the mechanical module, which will bring Agent i to a stop. In both cases, Agent j （$j < i$）will not be affected by Agent i and keep moving, and Agent k （$k > i$）will obstructed by Agent i.

$0 < v_1^k < \varepsilon_1$ represents that the velocity of Agent i is below the normal range. Leader and other agents will not able to reduce the speed, and Agent i will obstruct the motion of other agents. In this case, Agent i is considered to be faulty, and be the obstacle of other agents.

2) $\varepsilon_1 < v_1^k < \varepsilon_2$. In the case, although the velocity of Agent i is different from the normal one, but it still in the threshold value range. Since Agent i is one of the followers, Agent j ($j < i$) can continue moving at the normal speed, while Agent k ($k > i$) need to adjust the motion speed affected by Agent i. Then difference motion speed will caused the dismiss of flocking. So Agent i is set to be a obstacle in this paper.

3) $v_1^k < \varepsilon_2$ represents the velocity of Agent i is beyond the threshold value range. Other agents cannot follow Agent i to move in high speed. So, the Agent i is considered to be faulty. Owing to the high speed, Agent i will crash into the front Agents. And Agent i is set to be a obstacle for other agents in this case.

4 Fault agent diagnosis and avoidance

According to the above analysis, the fault agent avoidance method is given as follows.

1) When the leader has been faulty, the flocking will not be able to continue due to the loss of the leader. So the top priority is to select new leader in the remains. This paper uses the shortest distance approach to choose new leader to continue the motion of flocking, such as Agent 2 in Figure 2 as a new leader. After the selection of new leader, the follow strategy has to make a change correspondingly to maintain the shape of flocking, shown in Figure 2.
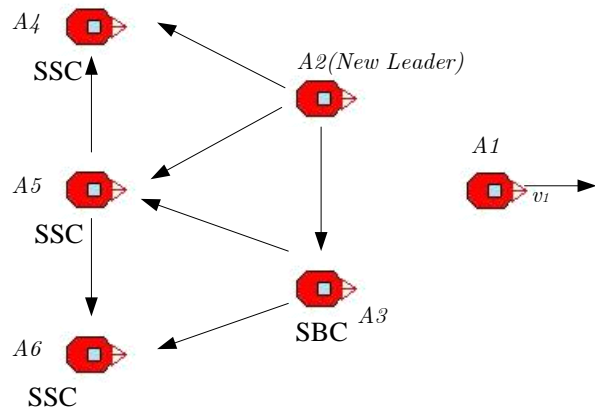
FIGURE 2 Flocking model of multi-agent with a new leader and a fault leader

Compared with Figure 1, Figure 2 shows a modification of the follow strategies on Agent 2 and Agent 3. Agent 4, 5 and 6 do not have any change. Shortly after the new leader and the new follow strategies have defined, the focus of flocking is to avoid Agent 1. There already have lots of literatures about the methods of avoiding obstacles. In terms of the complexity of routes of the flocking, different methods can apply.

2) When Agent i has malfunctions, it will affect the agents that have follow relations with Agent i. It is necessary to modify the follow strategy. Take an example of Agent 2 with malfunctions, as shown in Figure 3. As it can be seen from the comparison of Figures 1 and 3, the follow strategies of Agent 3 and Agent 4 need a quick modification when malfunctions occur in Agent 2. Agent 5 and 6 do not need a change of the strategies. When the new model of the flocking is defined, the flocking will start the avoidance process of Agent 2 by following the leader.
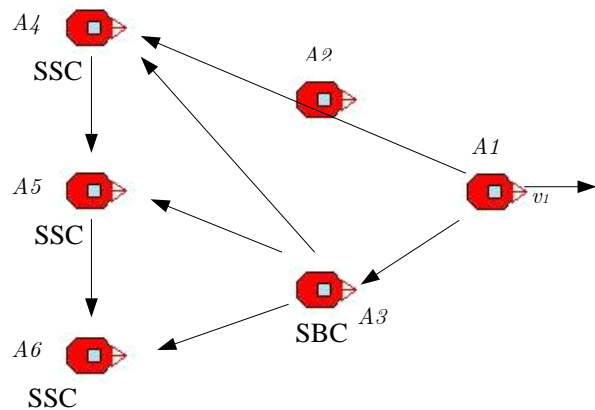
FIGURE 3 Flocking model with fault Agent 2

In summary, the algorithm of the fault agent diagnosis and avoidance for flocking of multi-agent is shown in Table 1.

TABLE 1 Fault diagnosis and obstacle avoidance algorithm of multi-agent

---

Initialization

Set the follow strategy as SBC $(A_1, A_2, l_{12}, \varphi_{12})$ and SCC $(A_i, A_j, A_k, l_{ij}, l_{ik})$ （i,j,k=2,3,4,5,6）

% Fault diagnosis and obstacle avoidance algorithm of Leader
for $k=1$ to maxsteps
       for i=1

           if $0 \le v_1^k < \varepsilon_1$

           Set leader(Agent 1) to be faulty and be a obstacle
           Choose Agent 2 to be the new leader

           Update follow strategy SBC $(A_2, A_3, l_{23}, \varphi_{23})$,SCC $(A_i, A_j, A_k, l_{ij}, l_{ik})$

           New leader leads followers to avoid the collision with the agent1.

       else if $\varepsilon_1 < v_1^k < \varepsilon_2$

           Update v_i for Agent i *(i=2,3,4,5,6)*

           else if $v_1^k > \varepsilon_2$

           Choose Agent 2 to be the new leader

           Update follow strategy SBC $(A_2, A_3, l_{23}, \varphi_{23})$,SCC $(A_i, A_j, A_k, l_{ij}, l_{ik})$

           New leader guides followers to move on
           end if

       end if
       end if
end
%Fault diagnosis and obstacle avoidance of followers
for i=2 to 6

       if $0 \le v_i^k < \varepsilon_1$

           Set Agent i to be faulty and be a obstacle
           if i=2

           Update follow strategy of agent i+1 as SBC $(A_1, A_{i+1}, l_{1(i+1)}, \varphi_{1(i+1)})$

       else update follow strategy of agent i+1 as SCC $(A_{i+1}, A_j, A_k, l_{(i+1)j}, l_{(i+1)k})$

       end if
       Other agent moves on with avoiding agent i

       else if $\varepsilon_1 < v_i^k < \varepsilon_2$

       Update v_i for Agent i *(i=2,3,4,5,6)*

       else if $v_i^k > \varepsilon_2$

           Set Agent i to be faulty and be a obstacle
           if i=2

           Update follow strategy of agent i+1 as SBC $(A_1, A_{i+1}, l_{1(i+1)}, \varphi_{1(i+1)})$

           else update follow strategy of agent i+1 as SCC $(A_{i+1}, A_j, A_k, l_{(i+1)j}, l_{(i+1)k})$

           end if
       Other agent moves on with avoiding agent i
       end if
       end if
       end if
end

---

5 Simulation and results

The simulation model of 6 amigo mobile robots is prepared using the Matlab/Simulink environment. Each robot is regarded as an agent. It is assumed that the multi robots move with uniform velocity. Under normal condition, the multi robots in the flocking can move with uniform velocity until reaching the destination. Considering two flocking motion modes, one is linear motion, the other is linear motion with 45 degrees. The desired trajectory is shown in Figure 4.

a) linear motion of flocking

b) linear motion with 45 degrees of flocking

FIGURE 4 The desired trajectory of the motion of flocking
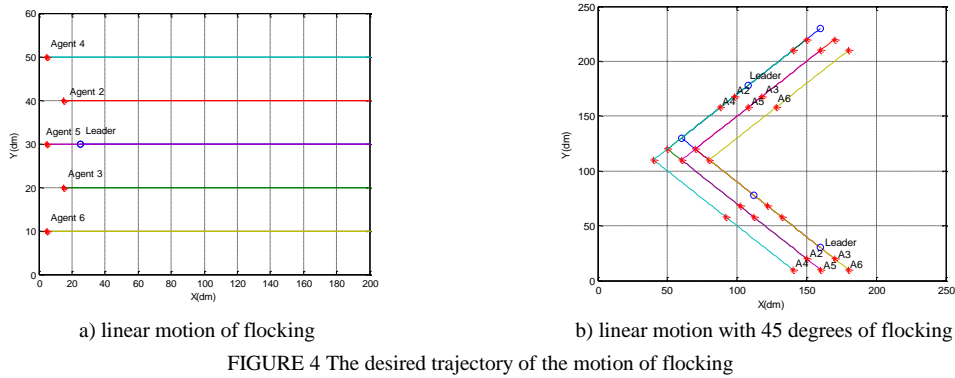
5.1 SIMULATION EXPERIMENT 1: LEADER FAULT

The original positions of A1~A6 are [25,30], [15,40], [15,20], [5,50], [5,30], [5,10]. The target of the motion of flocking is to move 100s. Leader is set to stop at t=10s.

According to the fault agent diagnosis and avoidance algorithm shown in Table 1, Agent 2 is set to be new leader to guide other agents. The obstacle avoidance of flocking is shown in Figure 5.

a) Original position when t=0s

b) Leader is faulty when t=10s

c) Agent 2 lead the motion of flocking when t=25s

d) Locations of all the agents when t=30s

e) Locations of all the agents when t=40s

f) Locations of all the agents when t=100s
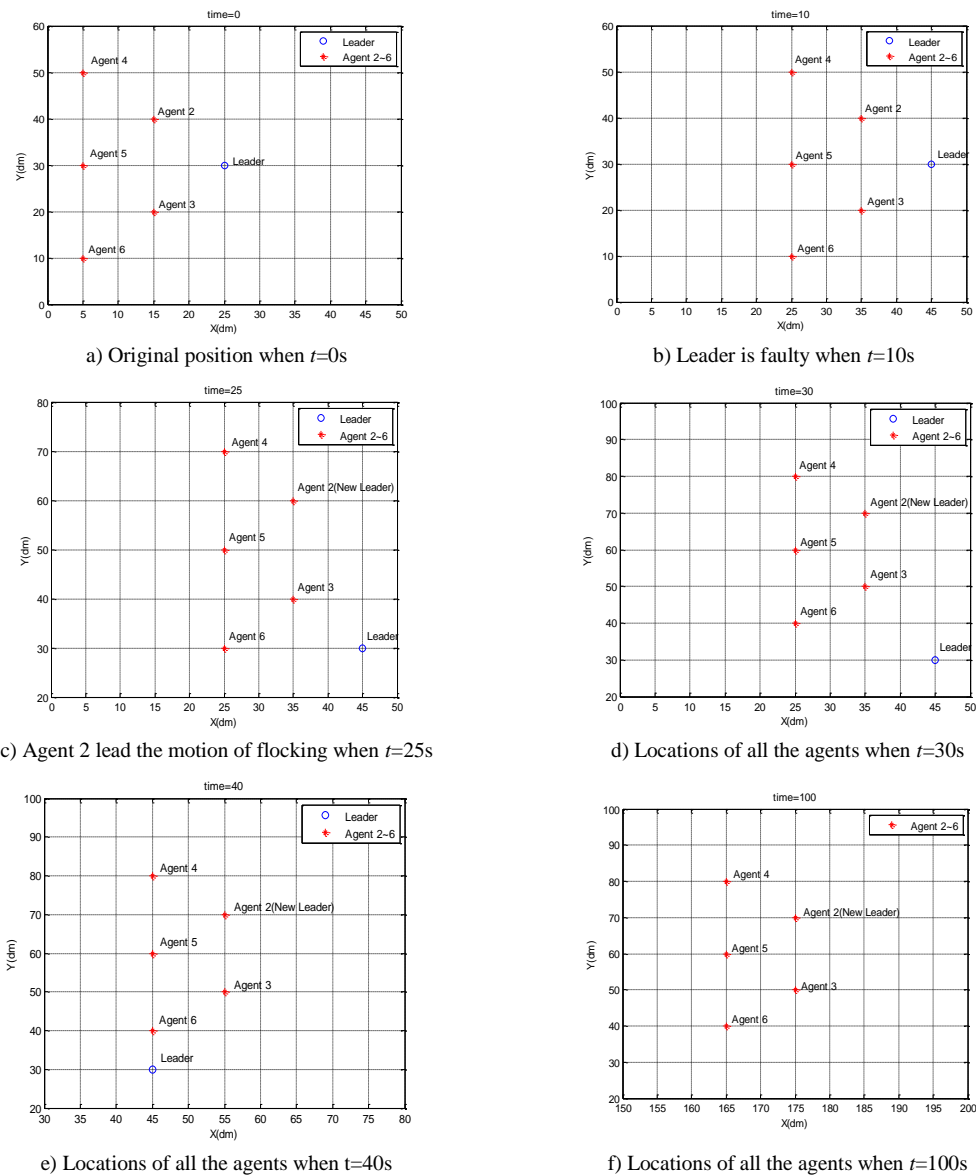
FIGURE 5 The motion of flocking with fault leader

113

As shown in Figure 5a shows the original position of 6 Agents. In Figure 5b, the location of the leader being faulty is marked at the time of *t* = 10s. The failure is simulated through setting the velocity of leader as zero. In Figures 5c and 5d, the location of other agents are marked when *t* = 25s and *t* = 30s respectively. Detecting the fault leader, Agent 2~Agent 6 choose Agent 2 as the new leader. And the new leader will guide other agents to avoid the fault leader. Figures 5e and 5f present the location of Agent 2~Agent 6. The results shows that the flocking of multi-agent can re-choose new leader quickly, and can continue the motion with avoiding the fault leader under the guidance of new leader.

5.2 SIMULATION EXPERIMENT 2: AGENT 2 FAULT

In this experiment, the flocking moves along a linear motion with 45 degrees. The normal motion trajectory is shown in Figure 4b. The original positions of Agent1 to Agent 6 are [160, 30], [150, 40], [170, 20], [140, 50], [160, 30], [180, 10]. The requirement of flocking is that leader should turn at the position of Y=130 and move along a diagonal direction, until to the position of X=160.

Agent 2 is set to be faulty when *t*=15s. Its velocity decreases. The results of avoidance of Agent 2 are shown in Figure 6.

a) The locations of all agents when *t*=15s

b) The locations of all agents when *t* = 35s

c) The locations of all agents when *t*=50s

d) The whole trajectory of the leader
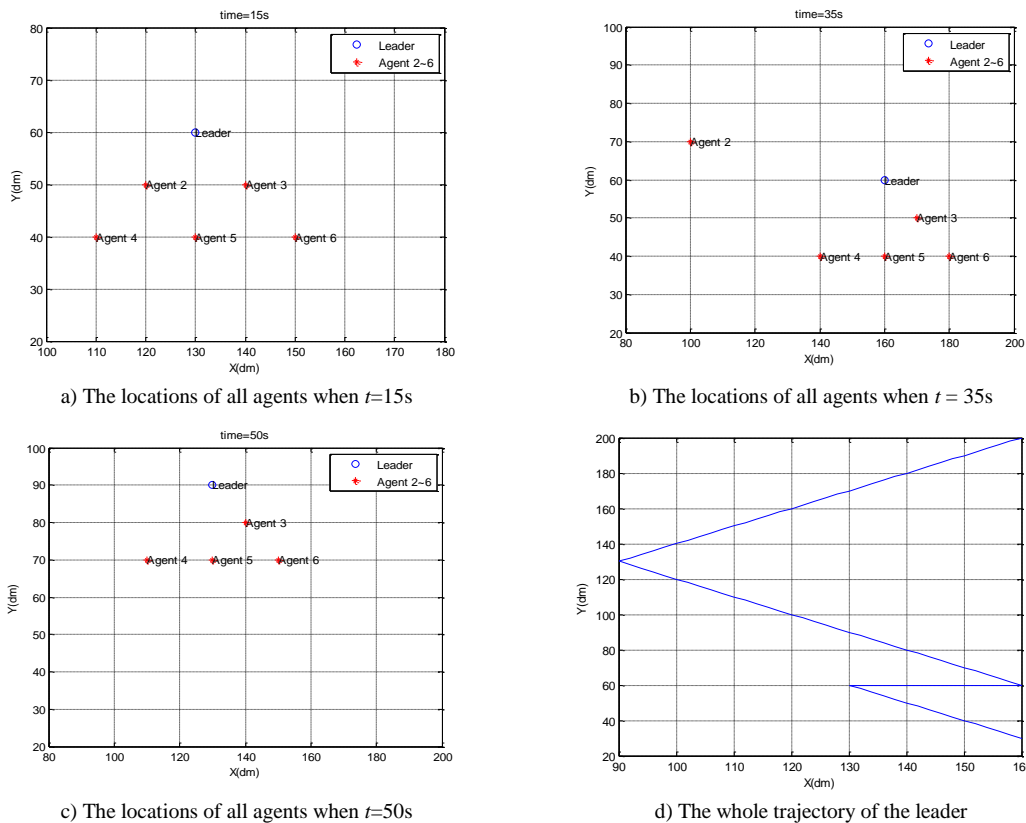
FIGURE 6 The motion of flocking with fault agent 2

Figure 6 shows that the flocking avoids the Agent 2 successfully and accomplishes the task. In Figure 6a, the location of the Agent 2 being faulty is marked when *t* = 15s. The speed of Agent 2 is assumed to be 1 dm/s in the experiment. Figure 6b shows the locations of all agents when *t* = 35s. Additionally, the agents successfully avoid the malfunction agent and start to move along a diagonal direction. Figure 6c shows the locations when *t* = 50s. Figure 6d presents the whole trajectory of the leader. It shows that when *t* = 15s, the leader is at the location of [130, 60]. At the same time, Agent 2 becomes faulty. The leader cannot move along the diagonal direction, as Agent 5 will collide with Agent 2. As a result, the leader starts to modify the motion direction to avoid the obstacles, and will not resume their original diagonal-direction motion until the completion of the obstacle avoidance.

6 Conclusions

In this paper, we introduced the motion of flocking of multi-agent with a leader. According to the velocity of agent, we classified the fault agent into six types, three for leader and three for follower agents. The fault agent diagnosis and avoidance is implemented by the method proposed in this paper. The simulation results confirm that the proposed method has good fault diagnosis and obstacle avoidance effect.

In reality, more agents will be faulty and more complicated fault types will occur in the motion of flocking. We plan to further investigate more fault types in flocking of multi-agent. Also we plan to study different algorithm to improve the fault agent avoidance efficiency.

References

[1] Tan M, Wang S 2013 Research progress on robotics *Acta Automatica sinica* **39**(7) 963-72

[2] Yang Y, Souissi S, Defago X 2011 Fault-tolerant flocking for a group of autonomous mobile robots *The Journal of Systems and Software* **84**(1) 29-36

[3] Suissi S 2007 Fault resilient cooperation of Autonomous Mobile robots with unreliable compass sensors *JAIST Japan*

[4] Wander A, Forstner R 2013 Innovative fault detection, isolation and recovery on-board spacecraft: study and implementation using cognitive automation *Control and Fault-Tolerant Systems* 336-41

[5] Canepa D, Gradinaiu M 2007 Stabilizing flocking via leader election in robot networks *INRIA France* 52-66

[6] Guifen S, Huajing F 2005 Fault tolerance of multi-robot formation based on adjacency matrix *Huazhong Univ. of Sci.& Tech. (Nature Science Edition)* **33**(3) 39-42

[7] Naixue X, Jing H, Yan Y, et 2010 A Survey on decentralized flocking schemes for a set of autonomous mobile robots *Journal of communications* **5**(1) 31-7

[8] Martin A, Emami M R 2013 A fault-tolerant approach to robot teams.*Robotics and Autonomous Systems* **61**(12) 1360-78

[9] Hui Y, Yongji W, Lei C 2005 Control of stable flocking motion of multipy-agent with a leader .*Huazhong Univ. of Sci.& Tech.(Nature Science Edition)(in Chinese)* **33**(8) 56-8

[10] Olfati-Saber R 2006 Flocking for multi-agent dynamic systems: Algorithms and Theory *IEEE Transactions on Automatic control* **51**(3) 401-20

[11] Moshtagh N, Jadbabaie A, Daniilidis K 2005 Vision-based distributed coordination and flocking of multi-agent systems *Robotics: Science and Systems* 41-8

[12] Antonelli G, Arrichiello F, Chiaverini S 2010 Flocking for multi-robot systems via the Null-Space-based Behavioral control *Swarm Intelligent* **4** 37-56

[13] Su H, Wang X, Lin Z 2009 *IEEE transactions on automatic control* **54**(2) 293-307

[14] Daige M, Koutsoukos X 2007 *IEEE Transactions on Robotics* **23**(2) 353-69

[15] Deshpande P, Menon P, Edwards C 2011 Formation control of multi-agent systems with double integrator dynamics using delayed static output feedback *IEEE Conference on Decision and Control and European Control Conference* 11 3446-51

[16] Desai J P, Ostrowski I, Kumar V 1998 Controlling formation of multiple mobile robots *IEEE International Conference on Robotics and Automation* 2864-9

Authors

Liu Yutian, born on August 8, 1980, Zhejiang, China

Current position, grades: lecturer Faculty of Electronic and Information Engineering, Zhejiang Wanli University.
University studies: PhD on control theory and control engineering at Zhejiang University.
Scientific interest: obstacle avoidance, fault diagnosis and intelligent control.
Publications: 12 papers.

Hu Junjie, born on October 16, 1965, Ningbo, China

Current position, grades: lecturer, Faculty of Electronic and Information Engineering, Zhejiang Wanli University.
University studies: M.S. degree on computer and application at Hangzhou Dianzi University.
Scientific interest: intelligent control, nonlinear system, and computer network.
Publications: 5 papers.

Computer and Information Technologies

Quantum public-key cryptosystem without quantum channel between any two users based on the Bell state measurement

Xiaoyu Li[*], Dai Wang

School of Information Engineering, Zhengzhou University, Zhengzhou City, China

Abstract

In this paper, a quantum public-key cryptosystem without quantum channel between any two users based on the Bell state measurement is presented. A user Alice shares a set of Einstein-Podolsky-Rosen (EPR) pairs with key management centre (KMC) as the private key and the public key. By performing the Bell state measurement on the public key and the auxiliary qubits any other user can send encrypted message to Alice. On the other hand, digital signature can also be achieved by this public-key cryptosystem. The laws of quantum physics guarantee the unconditional security of this public-key cryptosystem. No quantum channels are needed between any two users. So it is easier to carry out in practice and more robust against possible attacks.

Keywords: public-key, quantum cryptography, EPR pair, the Bell state measurement, digital signature

1 Introduction

In cryptography, people integrate the original information (called "the plain text") with some auxiliary information (called "the key") to produced encrypted information (called "the cipher text") by a cryptographic algorithm. Anyone who has not the key cannot recover the plain text from the cipher text. So the cipher text can be transmitted though an unsecure channel without any danger to leak the plaintext. Two users can fulfil secret communications through an insecure channel as long as they have shared the key before. As a result key distribution becomes the most important and difficult problem for people to complete secret communication. In fact, there are nearly no unconditionally secure classical key distribution protocols in classical cryptography.

Quantum key distribution (QKD) protocol is a good solution to this problem. In QKD protocols people can achieve unconditional security using the special physical properties of quantum system. C. H. Bennett and G. Brassard provided the first quantum key distribution protocol [1] in 1984. Since then people have developed many quantum key distribution protocols, such as the EPR protocol [2], B92 protocol [3], Lo-Chau protocol [4], et al [5-10]. Experimental work for OKD has also been finished in laboratory. Bennett, Bessette and Brassard first realized BB84 protocol in 1992 [11]. Now QKD protocol in optical fibre has been completed beyond 150 km [12] while QKD protocol in free space has also been achieved over a distance of 1 km [13].

Traditional QKD protocols belong to symmetrical key protocols. But all symmetrical key protocols are faced with a serious problem: how to distribute and manage keys if there are a lot of users in the cryptosystem? If there are N

users, which need to communicate with each other, every user must share a key with any one of the other users. So every user should keep N-1 keys secret so that no one can steal them. At the same time, every user should complete key distribution with any one of the other N-1 user. Obviously, it is an arduous task when N is a large number! Furthermore, in practice maybe the users do not trust each other so that it is impossible for them to perform key distribution. It is known that public-key cryptosystem can overcome this difficulty in classical cryptography, such as RSA algorithm [14]. In public-key cryptosystem a user has (public key, private key) pair in which the public key and the private key cannot be deduced from each other. The private key is used to decrypt the message encrypted by the public key while the public key is used to decrypt the message encrypted by the private key. Every user keeps his private key secret so as that no one can get it. At the same time, a key management centre keeps all users' public keys, which are open to everyone. If a user Bob wants to send a secret message to another user Alice, he first asks KMC for Alice's public key and encrypts the message by the public key to get the cipher text. Then Bob sends the cipher text to Bob. When Bob receives the cipher text, he can decrypt it by his private key to recover the original message. Any eavesdropper who catches the cipher text cannot recover the original message because he or she does not hold Bob's private key. Public-key cryptography technology has become one of the most important tools to safeguard information security in modern society, such as commercial affairs, military affairs, network communications et al. But Peter Shor proved that RSA algorithm is unsafe on future quantum computer in 1994 [15]. So the classical public-key cryptosystem based on RSA algorithm will been crashed

---

[*] *Corresponding author* e-mail: iexyli@zzu.edu.cn

by attacks on quantum computer. Quantum public-key technology can provide a good alterative solution. In 2001 Gottesman first presented a quantum one-way function to design quantum digital signature protocol, which may be used in a public-key system. A similar scheme is provided in [17]. In 2008 Nikolopoulos put forward the first quantum public-key cryptosystem [18] based on the property of single-particle rotation of unknown quantum states which can provide unconditional security. Since then a few public-key protocols have been studied [19-22].

In this paper, we provide a quantum public-key cryptosystem based on the Bell state measurement. Users and KMC share EPR pairs as the public key and the private key. With the help of KMC, N users can communicate with each other securely. Moreover digital signature for message can be fulfilled naturally by the public-key cryptosystem. There are no quantum channels needed between any two users. So it's easy to carry out in practice. We prove that the cryptosystem is secure against possible attack.

2 Basic idea

As known a quantum two-state system is called a qubit. A qubit may be in one of the four possible states

$$|0>, |1>, |+>=\frac{1}{\sqrt{2}}(|0>+|1>), |->=\frac{1}{\sqrt{2}}(|0>-|1>). \quad (1)$$

Obviously, the four states are not orthogonal to each other. Therefore, it is impossible to determine in which state a qubit is with certainty. On the other hand, they form two complete orthogonal bases in which we can measure a qubit:

$$B_{01}=\{|0>, |1>\}, \quad B_{+-}=\{|+>, |->\}. \quad (2)$$

A two-qubit system can be in one of the four Bell states:

$$|\Phi^+>=\frac{1}{\sqrt{2}}(|00>+|11>), |\Phi^->=\frac{1}{\sqrt{2}}(|00>-|11>),$$

$$|\Psi^+>=\frac{1}{\sqrt{2}}(|01>+|10>), |\Psi^->=\frac{1}{\sqrt{2}}(|01>-|10>). \quad (3)$$

Such a two-qubit system is often called an EPR pair. It is easy to find that the four Bell states forms a complete orthogonal basic vector set in which people can measure a two-qubit system. Such measurement is called the Bell state measurement, which has been carried out [23].

We assume that Alice and Bob share M EPR pair in the state:

$$|\Phi^+>_{12}=\frac{1}{\sqrt{2}}(|0>_1|0>_2+|1>_1|1>_2), \quad (4)$$

in which qubit 1 is hold by Alice and qubit 2 is bold by Bob. Then to each EPR pair Bob creates an auxiliary (denoted as qubit B) in the state |0> and put it together with qubit 2. So the state of the whole three-qubit system is:

$$|S>=\frac{1}{\sqrt{2}}(|0>_1|0>_2|0>_B+|1>_1|1>_2|0>_B). \quad (5)$$

It can be rewritten as:

$$|S>=\frac{1}{\sqrt{2}}[|0>_1(|\Phi^+>_{2B}+|\Phi^->_{2B})+$$

$$|1>_1(|\Psi^+>_{2B}-|\Psi^->_{2B})]. \quad (6)$$

Now Alice measures qubit 1 in basis $\{|0>, |1>\}$ while Bob performs the Bell state measures on the composed system of qubit 2 and qubit B. It is easy to find that Alice's measurement results and Bob's measurement result are correlated in Table 1.

TABLE 1 Correlation of their measurement results

Alice's measurement result	Bob's measurement result
$|0>_1$	$|\Phi^+>_{2B}$
	$|\Phi^->_{2B}$
$|1>_1$	$|\Psi^+>_{2B}$
	$|\Psi^->_{2B}$

So it's possible to establish a public-key cryptosystem based on the result above.

Now let us consider a public-key cryptosystem, which includes a key management centre (KMC) and N users. First, we have the Key Rule.

Key Rule:

$$|0>\rightarrow0, |1>\rightarrow1, |+>\rightarrow0, |->\rightarrow1,$$

$$|\Phi^+>\rightarrow0, \quad |\Phi^->\rightarrow0, |\Psi^+>\rightarrow1, \quad |\Psi^->\rightarrow1. \quad (7)$$

A user, e.g. Alice, creates M EPR pairs in which every EPR pair is in the state:

$$|\Phi^+>_{12}=\frac{1}{\sqrt{2}}(|0>_1|0>_2+|1>_1|1>_2). \quad (8)$$

Then Alice shares the EPR pairs with KMC in which qubit 1 of the EPR pair is hold by Alice and qubit 2 is bold by KMC. So the qubit sequence hold by Alice denoted as Q^A is just Alice's private key while the qubit sequence hold by Bob denoted as Q^K is just Alice's public key. The public key is open to every user while Alice keeps her private key secret in order that no one except herself can get it. Now another user, such as Bob, wants to send a secret message to Alice. E.g., the message may be an n-bit string denoted as *P*, which is just the plain text. To encrypt the plain text, Bob asks KMC for Alice's public key Q^K. After getting Q^K, to every qubit in Q^K Bob creates an auxiliary qubit in the state |0> and performs the Bell state measurement on the two qubits. At the same time Bob records his measurement result according to the Key Rule. Finally, Bob gets an M-bit string S. On the other hand, Alice measures Q^A in B_{01} and records her measurement

results according to the Key Rule. Finally Alice also gets an M-bit string denoted as *S'*. Alice and Bob mutually choose t bits from *S'* and *S* in which $t = M - n$ and compare them. If there are too many disagreements, they abandon the intention of communications and turn back to the beginning. Alternatively, they can be sure that no errors or eavesdroppers existing. So Alice has an n-bit string denoted as *K'* while Bob has an n-bit string denoted as *K*. It is obvious that $K' = K$. Next Bob perform an XOR operation on *P* and *K* to get a new n-bit string *PS* in which:

$$PS = P \oplus K. \qquad (9)$$

Then Bob sends *PS* to Alice through the public classical channel. When Alice receives it, she performs an XOR operation on *PS* and *K'* to get *P'*

$$P' = PS \oplus K'. \qquad (10)$$

From (9) and (10) we get

$$P' = P. \qquad (11)$$

So Alice has gotten the plain text which Bob wants to send her. In section IV we will prove that no one except Alice and Bob can get the plain text. So Bob succeeds in sending a secret message to Alice. It is easy to find that Alice and Bob needn't exchange qubits at all. So no quantum channels are needed between them. This is a notable advantage of our public-key cryptosystem.

To guarantee the public-key work, there is still a problem, which must be solved. The public key and the private key, which are all parts of the M EPR pairs lose correlations after Alice's and Bob's measurements. So both the public key and the private key no longer exist after a communication process, or in other words, the (public key, private key) pair can be used for only one time. But many users may need to communicate with Alice and one user may need to send a lot of secret messages to Alice. So KMC and Alice should share L ($L \gg N$) (public key, private key) pairs. Each (public key, private key) pair of Alice should be given unique id number. So does every user in the public-key cryptosystem.

So we can design a quantum public-key cryptosystem based on the idea above.

3 Quantum public-key cryptosystem without quantum channels using the Bell state measurement

3.1 BUILDING THE PUBLIC-KEY CRYPTOSYSTEM

First, we assume that there are *N* users and a KMC in our public-key cryptosystem. There are a classical channel and a quantum channel. The classical channel is open so that everyone can listen to it and send classical information to others. However, the classical channel is authenticated so that everyone can assure the identity of the counterpart who is communicating with him. The quantum channel is insecure. Everyone can catch the qubit transmitted through it and send fake qubits to any other one without being found. Every user, such as Alice, creates L EPR pairs and

share them with KMC in which the first qubit (qubit 1) is hold by herself and the second qubit (qubit 2) is hold by KMC. So the Alice's public keys set is denoted as:

$$K_{PK} = \{ (i, Q_i^K), \quad i = 1, 2, ..., L \}, \qquad (12)$$

in which M-qubits sequence and i is the id number. On the other hand, Alice keeps her private keys denoted as

$$K_{PA} = \{ (i, Q_i^A), \quad i = 1, 2, ..., L \}. \qquad (13)$$

All users' public keys are open to everyone, that is to say, any person can get any public key of any other user from KMC. But one public key can only be given to one user because it will be consumed and no longer exist. It must be pointed out that every user must keep his or her private keys absolutely secret. Certainly one private key can also be used for one time.

3.2 PROCESS OF THE SECRET COMMUNICATION

If user Bob wants to send a secret message denoted as an n-bit string P to another user Alice, they perform the following steps:

Step 1: Bob asks KMC for one of Alice's public keys.

Step 2: KMC chooses a public key (j, Q_j^K) from Alice's K_{PK} at random and gives it to Bob through the quantum channel while KMC sends the id number j to Bob through the classical channel.

Step 3: After receiving (j, Q_j^K) and *j*, Bob sends *j* to Alice through the classical channel.

Step 4: After receiving the id number *j*, Alice queries it in her K_{PA} and gets the corresponding private key (j, Q_j^A) in order to decrypt the cipher text received.

Step 5 (error-checking): Alice chooses *t* qubits from (j, Q_j^A), where $t = M - n$. Then she measures them in basis B_{01} or B_{+-} at random and declares her choice of each measurement basis. Bob chooses the corresponding qubits in (j, Q_j^K) and measures them in the identical basis just as Alice. Finally Alice and Bob compare their measurement results. If there are too many disagreements, they abandon it and turn back to step 1. Or they continue into the next step.

Step 6: To each of the left *n* qubits (denoted as qubit 2) in (j, Q_j^K), Bob creates an auxiliary qubit (denoted as qubit B) and performs the Bell state measurement on the composed system of qubit 2 and qubit B. Bob records his measurement results according to the Key Rule. Finally, Bob gets an M-bit string *K*.

Step 7: To each of the left n qubit in (j, Q_j^A), Alice measures it in basis $\{|0>, |1>\}$ and records her measurement results according to the Key Rule. Finally Alice gets a string K'.

Step 8: Bob performs XOR operation on K and P to get the cipher text PS. Then Bob sends PS to Alice through the classical channel.

Step 9: When Alice receives PS, she performs XOR operation on PS and K' to get the decrypted text P'.

Obviously $P' = P$. So Alice has gotten the secret message which Bob wants to send her.

If Alice wants to send a secret message to Bob, they need only exchange the roles in the process above. So any two users can fulfil secret communications by this public-key cryptosystem.

3.3 DIGITAL SIGNATURE

First all users agree to the following rule:
Signature Rule:

$$|0> \rightarrow 0, \quad |1> \rightarrow 1 . \tag{14}$$

If Alice receives a secret message, which is claimed from Bob, how can she assure that it is really from Bob? Such problem can be solved by digital signature. Bob can sign the message to guarantee that it is just the message he wants to send Alice. Let us assume that Bob wants to send a string P to Alice. To produce the signed message, Bob performs as following steps:

Step 1: Bob produces an m-bit abstract PA of P using a hash algorithm, for example, SHA-1 algorithm.

Step 2: Bob chooses one of his private keys at random, such as (k, R_k^A). Then he performs measurement on the first m qubits of (k, R_k^A) in basis $\{|0>, |1>\}$ and records his results according to the Signature Rule. Finally, Bob gets an m-bit string PK.

Step 3: Bob performs an operation $PA \oplus PK$. Finally, he gets an m-bit string PSD, which is just the signed message.

Step 4: Bob attaches PSD and the id number k with the message P. So he gets a string PX which is just the plain text to be submitted to Alice.

Notice that the length of PX should be n. So the length of the original message P added with the length of k should be $n - m$. If P cannot satisfy it, we can always make it by dividing it into several parts or adding supplementary bits to it.

Now Bob and Alice can finish the communication as the steps in section III.

After Alice gets the plain text PX, she extracts the original message P, the signed message PSD and the id number k. To verify the signature, she performs the following steps.

Step 1: Alice asks KMC for Bob's k public key (k, R_K^K).

Step 2: Alice measure the first m qubits of (k, R_K^K) in basis $\{|0>, |1>\}$ and records her measurement results according to the Signature Rule. Finally she also gets an m-string PK' which is just equals to PK.

Step 3: Alice performs an operation $PSD \oplus PK'$. Finally she gets an m-bit string PA'.

Step 4: Alice produces the abstract PA of P by SHA-1 algorithm just as Bob does.

Step 5: Alice compares PA' and PA. If they are identical, the verification passes. Alice can be sure that the message is just from Bob.

4 Security of the public-key cryptosystem

This quantum public-key cryptosystem is secure. Two users can communicate with each other secretly. Any other people including KMC cannot get the message. We prove it as follows.

First, we assume that an eavesdropper, e.g., Eve, wants to get the message sent from Bob to Alice.

4.1 IMPOSSIBILITY FOR EAVESDROPPER TO GET THE MESSAGE

Eve can listen to both the classical channel and the quantum channel, trying to get the secret message from Bob to Alice. She can get the cipher text PS sent form Bob to Alice in step 5. At the identical time she also knows that the plain text is encrypted by Alice's j public key (j, Q_j^K).

As known PS is produced from $P \oplus K$. It's easy to deduce that $P = PS \oplus K$. Since Bob has gotten PS, he can get P as long as he gets K. But K is kept secret by Bob so that he won't give K to anyone. What Eve can do is to monitor the process that Bob creates the cipher text, trying to get some information about K. First Eve can listen to all the classical information exchanged between Alice and Bob. But Alice' K' or Bob's K is from the measurement results on (j, Q_j^A) or (j, Q_j^K), which Eve doesn't posses at all. So Eve can get no information about even one bit of K or K'.

Second Eve may catch (j, Q_j^K) when KMC sends it to Bob in step 2. But she can't measure (j, Q_j^K) because at present (j, Q_j^K) contains no information about K which is produced by the random measurement results of Bob in step 5. If Eve measures (j, Q_j^K) now, she can only get a random string, which has nothing about K. Furthermore, if Eve measures (j, Q_j^K), the qubits will collapse into eigenstates and no longer entangle with the qubits in (j, Q_j^A). So Alice's measurement results on (j, Q_j^A) will have no correlations with Bob's measurements results on (j, Q_j^K) at all, that is to say, Alice and Bob will be sure to find too many disagreements between K or K' in step 5 and abandon the process of communication. Eve can get nothing by this attack method. The probability that Alice and Bob just get the identical value of the t bits, or in other words, Eve succeeds in getting K, is

$$P_{error} = \left(\frac{1}{2}\right)^{t}. \tag{15}$$

If $t = 200$, we have

$$P_{error} = \left(\frac{1}{2}\right)^{200} \approx 10^{-60}. \tag{16}$$

It is a number too small to imagine. So Eve's attack is sure to fail.

On the other hand, Eve may take the strategy of entanglement attack. First Eve catches (j, Q_j^K) in step 2. Then to each qubit (denoted as qubit 2) in (j, Q_j^K), she creates an auxiliary qubit (denoted as qubit E) and performs CNOT operation on qubit 2 and qubit E in which qubit 2 is the control qubit and qubit E is the target qubit. So qubit E is also entangled with the EPR pair (qubit 1, qubit 2). Then Eve tries to get K from the correlated collapse of qubit E and qubit 2. It's easy to prove that such strategy can't succeed. After Eve's CNOT operation, the state of the whole three-qubit system turns into:

$$|T> = \frac{1}{\sqrt{2}}(|0>_1|0>_2|0>_E + |1>_1|1>_2|1>_E). \tag{17}$$

Then in step 5 Alice and Bob perform error-checking. They measures qubit 1 and qubit 2 in the identical basis. If the basis is B_{01}, Alice and Bob will get the identical result. So Eve escapes from being found by Alice and Bob. On the other hand, Equation (17) can be rewritten as:

$$|S> = \frac{1}{2\sqrt{2}}[(|+>_1|+>_2 + |+>_1|->_2 + \\ |->_1|+>_2 + |->_1|->_2)|0>_E + \\ (|+>_1|+>_2 - |+>_1|->_2 + \\ |->_1|+>_2 + |->_1|->_2)|1>_E]. \tag{18}$$

If the basis is B_{+-}, the probability that Alice and Bob get the identical result is 1/2. Since Alice chooses the basis B_{01} or B_{+-} at random, the average probability that Alice and Bob get the identical probability for one qubit is:

$$p = \frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}. \tag{19}$$

So the probability that all the measurement results for the t qubits, or in other words, the probability that Eve escapes from being found is:

$$P_{error} = \left(\frac{3}{4}\right)^{t}. \tag{20}$$

If $t=200$:

$$P_{error} = \left(\frac{3}{4}\right)^{200} \approx 10^{-25}. \tag{21}$$

It is a very small probability, which can be ignored. So the strategy of entanglement attack is also invalid.

4.2 IMPOSSIBILITY FOR KMC TO GET THE MESSAGE

Just like Eve, KMC cannot get the message that Bob sends to Alice although it keeps the public keys and joins in the communications process. KMC cannot measure (j, Q_j^K) and cannot perform attack of entanglement because it can do nothing more than Eve can do. We have proved that such attacks cannot succeed.

On the other hand, KMC may also take a complex strategy of attack. To Alice's public key (j, Q_j^K), KMC creates M EPR pair and split them into two M-qubit sequence: (j, FQ_j^K) and (j, FQ_j^A). When Bob asks for Alice's public key, it gives (j, FQ_j^K) to Bob. Then KMC measures (j, FQ_j^A) while Bob measures (j, FQ_j^K). On the other hand KMC measures (j, Q_j^K) while Alice measures (j, Q_j^A). KMC tries to get some information about K or K' by this method. Obviously Bob's measurement results on (j, FQ_j^K) and KMC's measurement results on (j, FQ_j^A) are identical. KMC's measurement results on (j, Q_j^K) and Alice's measurement results on (j, Q_j^A) are identical. However in step 5 Alice and Bob perform error-checking in which they compare Alice's measurement results on (j, Q_j^A) and Bob' measurement results on (j, FQ_j^K). It is easy to find that there are no correlations between Alice's results and Bob's results because (j, Q_j^A) and (j, FQ_j^K) are not entangled with each other. So the probability that Alice and Bob get the identical measurement result for one qubit is 1/2. Alice and Bob measure t qubits respectively in step 5. So the probability that they just get identical results for all the t qubits is:

$$P_{error} = \left(\frac{1}{2}\right)^{t}. \tag{22}$$

If $t = 200$:

$$P_{error} = \left(\frac{1}{2}\right)^{200} \approx 10^{-60}. \tag{23}$$

So we can conclude that Alice and Bob are sure to find something wrong and abandon the process of

communication, that is to say, KMC can't succeeds in getting the secret messages.

4.3 SECURITY AGAINST FAKE MESSAGE ATTACK FROM EVE

Since Eve cannot get the secret message, can she make Alice to receive a fake message? We prove that it is impossible. Eve may try to catch the cipher text *PS* from Bob to Alice and produce a fake message to send to Alice. But Alice will perform XOR operation on PS and K' to recover the plain text. Although Eve can send Alice any fake cipher text, she can never make Alice to get the message that she wants Alice to accept because she does not hold *K* (or *K'*). Whatever Eve does, the probability that she make Alice to accept a specified message is no more than the probability that she guess all the bits of *K* correctly, which is

$$P_{error} = \left(\frac{1}{2}\right)^n . \tag{24}$$

If *n* = 1000:

$$P_{error} = \left(\frac{1}{2}\right)^{1000} \approx 10^{-300} . \tag{25}$$

That is to say, such attack also fails.

On the other hand, Eve may catch Alice's public key (j, Q_j^K) when it is transmitted from KMC to Bob. Then she sends Bob a fake key (j, FQ_j^K) with the intention to make Bob to get a fake string PK to encrypt the plain text. But Alice and Bob performs error-checking in step 5. Since (j, FQ_j^K) isn't entangled with (j, Q_j^A), Bob's measurement results on (j, FQ_j^K) have no correlations with Alice's measurement results on (j, Q_j^A). The probability they just get the identical result for one qubit is 1/2. So the probability that they get the identical results for all the *t* qubits is:

$$P_{error} = \left(\frac{1}{2}\right)^t . \tag{26}$$

If *t* = 200:

$$P_{error} = \left(\frac{1}{2}\right)^{200} \approx 10^{-60} . \tag{27}$$

So Eve still fails.

4.4 SECURITY AGAINST FAKE MESSAGE ATTACK FROM KMC

It is easy to prove that KMC cannot make Alice to get a fake message, either. KMC can catch PS and send a fake

cipher text to Alice just as Eve. On the other hand, KMC can give Bob a fake key (j, FQ_j^K) just as Eve. Obviously, what KMC can do is no more than what Eve can do. We have proved that Eve cannot make Alice to accept his fake message. So KMC cannot make it, either.

4.5 SECURITY OF DIGITAL SIGNATURE

Now we prove that our cryptosystem can solve digital signature problem. Alice can affirm that the message she receives is really from Bob and the message is integral without being tempered. After Alice receives the cipher text, she decrypts it and extracts the original message *P*, the signed message PS and the id number k. Then Alice verifies the signature. First she asks KMC for Bob's *k* public key (k, R_k^K) and measures the first m qubits to get a string *PK'*. Then she performs XOR operation to recover *PA'*, where:

$$PA' = PS \oplus PK' . \tag{28}$$

On the other hand, Alice produces the abstract of *P* by SHA-1 algorithm. Finally, she gets *PA*. Notice that

$$PS = PA \oplus PK . \tag{29}$$

If Alice find *PA* = *PA'*, she gets

$$PK = PK' . \tag{30}$$

Such fact shows that the one who sends the message to Alice should be able to get a string identical to *PK'* which is produced by Alice's measurement result on (k, R_k^K). For a man who hasn't hold Bob's *k* private key (k, R_k^A), the probability that he just guess all the m bits of *PK'* is:

$$P_{error} = \left(\frac{1}{2}\right)^m . \tag{31}$$

If *m* = 100:

$$P_{error} = \left(\frac{1}{2}\right)^{100} \approx 10^{-30} . \tag{32}$$

It is an extremely small probability. So Alice can assure that the one must have Bob' *k* private key (k, R_k^A), or in other words, the one must be Bob. On the other hand, SHA-1 algorithm guarantees that any string except *P* cannot produce the abstract PA. The message *P* must be integral and unchanged. So this public-key cryptosystem provides a reliable digital signature method.

4.6 SECURITY AGAINST FORWARD SEARCH ATTACK

The forward search attack is a serous danger to classical public-key cryptosystems. Since Alice's public key is kept open by KMC, everyone can ask KMC for it. So Eve may

encrypt a large number of plain texts by Alice's public key to produce the same number of cipher texts and save them in her database. Then Eve catches every cipher text sent to Alice and queries them in her database. If she just finds that a cipher text, which she catches, is identical to one cipher text in her database, she can affirm that the corresponding plain text in her database is just the secret message sent to Alice. So Eve gets the secret message without being found. However, in our quantum public-key cryptosystem, the forward search attack is invalid because KMC keeps many public-key for Alice in which one public-key can be used for one time. Two cipher texts, which are produced from the identical plain texts, are completely different. Eve's database is useless. She can never find the correct plain text in her database from a cipher text, which she catches.

So the forward search attack is sure to fail. This is a big advantage of this quantum public-key system.

4.7 SECURITY AGAINST RESEND ATTACK

In classical public-key cryptosystem, Eve may perform resend attack. She can catch a message sent from Bob to Alice and make a copy of it. After some time she resends the message to Bob again. When Alice receives the cipher text, she can decrypt it to the plain text without finding anything wrong. Therefore, Eve makes Alice to accept an outdated and repeated message although Eve completely knows nothing about the message at all. In classical public-key cryptosystem, to defeat resend attack people should add a timestamp to the original plain text so as Alice can find that the message is repeated. But to produce and verify the timestamp user need pay more cost.

In this quantum public-key cryptosystem, resend attack is also invalid. First Eve can catch PS and make a copy of it when it is sent from Bob to Alice. But resending it is pointless. The public key and private key to be used to encrypt and decrypt the original message have been consumed. When Eve wants to send Alice a message, they must perform all the steps in the communication process. So Alice will ask Eve for the id number for the key. However, the private (j, Q_j^A) no longer exists. If Eve still gives Alice the id number j, Alice will find that this is a resend attack at once. If Eve provides another id number, such as j', in step 5 (error-checking) Alice will perform measurement on t qubits of $(j', Q_{j'}^A)$. Then Eve must provide her measurement results of $(j', Q_{j'}^K)$ which must be identical to Alice's measurement results. If Eve can't provide them, she is found by Alice right away. Even if Eve has gotten $(j', Q_{j'}^K)$ from KMC, she can't succeed, either. In step 7 Alice gets a string NK', which she will use it to decrypt the cipher text which she receives. But the string NK' has nothing to with the string K', which Bob used to encrypt the plain text. When Alice gets PS which is resent by Eve, she tries to decrypt PS with NK'.

Obviously, Alice can get nothing but a garbled string so that she can be sure this message is unreal. So Eve has no chance to resend the repeated message to Alice at all.

4.8 SECURITY AGAINST CHOSEN PLAIN TEXT ATTACK

In a chosen plain text attack, Eve can obtain a random number of (plain text, cipher text) pairs of her choices, or in other words, she can get random cipher text for a specified plain text. Then Eve tries to find some information about the key by analyze the (plain text, cipher text) pairs. Chosen plain text attack is a power tool to crash many classical cryptographic algorithms if the number is large enough. But in our public-key cryptosystem one public key can be used for only one time. Any two cipher texts are produced by two different public keys. Furthermore, two identical plain texts are sure to be converted to completely different cipher texts. So Eve can find no laws which can help her to get something information about the key. In fact no matter how many (plain text, cipher text) pairs Eve may get, she can get nothing helpful to break the public-key cryptosystem. So the chosen plain text attack is impossible to succeed.

Now we have proved that our public-key cryptosystem is unconditionally secure.

5 Feasibility analysis of the public-key cryptosystem

First in this quantum public-key cryptosystem all that users need to do are performing the Bell state measurement on an EPR pair, performing single-particle measurements on a qubit and transmitting qubits through a quantum channel. All these have been realized in laboratory for many years. There are no unsolved technical difficulties at all. So it is easier to carry out in practice. On the other hand any two users need only to exchange classical information through a public classical channel. They don't need exchanging qubits at all. There are no quantum channels needed to connecting two users, which reduces the resource requirements and technical complexity greatly. So it is a big advantage of our quantum public-key cryptosystem.

Second, it is known that quantum cryptographic technology depends on the special physical properties of quantum systems. But quantum systems will inevitably undergo decoherence as time goes on. Once decoherence happens, it makes quantum to lose quantum coherence and to turn into classical systems so that all quantum cryptographic technology lose effectiveness. It's the most important threat to quantum cryptographic technology. In traditional quantum cryptographic protocols, such as quantum key distribution, we can complete the protocol as soon as possible before decoherence occurs. But in public-key cryptosystem, KMC needs to keep every user's public keys for a relative long time until a user asks for them. So decoherence is a problem which can not be avoided. To solve this problem, we can use the quantum system which has bigger time length of decoherence, for example,

photon in resonator. Another method is to make users to update their public keys periodically before decoherence happens. Using such methods this cryptosystem can work well to satisfy all users.

Third, in the public-key cryptosystem there is a public classical channel and a quantum channel needed. We have proved that eavesdroppers cannot get the secret message without being found. How about noise in these channels? Does it make the process of communication to error even fail? First, let us consider noisy classical channel. In step 3, Bob sends the id number j to Alice, which is necessary to go into the next step. If there are errors in transmission which cause a mistaken number j', communication between Bob and Alice is sure to fail. Likewise, in step 5 Alice and Bob need to exchange classical information for error-checking. The noisy information may cause mistaken results. Fortunately, classical error-correcting coding technology has been mature and powerful to deal with noisy channel. By error-correcting coding technology, it is easy to guarantee that classical information is transmitted with a very low error rate. So noise in classical channel doesn't affect our quantum public-key cryptosystem. On the other hand, in step 2 KMC sends Alice's public key (j, Q_j^K) to Bob through the quantum channel. If there are random errors existing, Bob will get mistaken bits in step 5 and step 6, which also cause communication to fail. The solution is quantum error-correcting coding technology. Although quantum error-correcting coding technology still cannot work as well as classical error-correcting coding technology, it is sufficient to accomplish reliable communications through most quantum channels.

6 Discussions and conclusion

In this quantum, public-key cryptosystem a public key can be used for only one time, which is a limit to the cryptosystem's capacity for work. If a user's public keys have been used up, no one can send message to him again. Developing cryptosystem with reusable public key is a solution to this problem. Replenish public keys periodically is another solution. We will discuss them in future work.

In this paper, we provide a quantum public-key cryptosystem without quantum channels between any two users based the Bell state measurement. Users use EPR pairs as public key and private key. The laws of quantum physics guarantee that no one except the two parts involved in communication can get the secret message. So it's unconditionally secure. At the same time the integrality and truth of the message exchanged can be verified by digital signature. There are no quantum channels needed between any two users. So it is easier to carry out in practice.

References

[1] Bennett C H, Brassard G 1984 *Proceedings of IEEE International conference on Computers Systems and Signal Processing* 175-9
[2] Ekert A K 1991 *Physical Review Letters* **67**(6) 661-3
[3] Bennett C H, Brassard G, Mermin N D 1992 *Physical Review Letters* **68**(5) 557-9
[4] Lo H K, Chau H F 1999 *Science* **283**(5410) 2050-6
[5] Nguyen T, Sfaxi M A, Ghernaouti-Hélie S 2006 *Journal of Networks* **1**(5) 9-20
[6] Qi B, Zhao Y, Ma X F, Lo H K, Qian L 2007 *Physical Review A* **75**(5) 052304
[7] Matsumoto R 2007 *Physical Review A* **76**(6) 062316
[8] Zhao Y, Qi B, Lo H K 2008 *Physical Review A* **77**(**5**) 052327
[9] Horodecki K, Horodecki M, Horodecki P, Leung D, Oppenheim J 2008 *IEEE Transaction on Information Theory* **54**(6) 2604-20
[10] Barrett J, Colbeck R, Kent A 2012 *Physical Review A* **86** 062326.
[11] Bennett C H, Bessette F, Brassard G, Salvail L, Smolin J 1992 *Journal of Cryptology* **5**(1) 3-28
[12] Kimura T, Nambu Y, Hatanaka T, Tomita A, Kosaka H, Nakamura K 2004 *Japanese Journal of Applied Physics* **43** L1109

[13] Buttler W T, Hughes, R J, Kwiat P G, Lamoreaux SK, Luther G G, Morgan G L, Nordholt J E, Peterson C G, Simmons C M 1998 *Physical Review Letters* **81**(15) 3283-6
[14] Rivest R, Sharmir A, Adleman L 1978 **21**(**2**) 120-6
[15] Shor P W 1994 *Proceedings of 35th Annual IEEE Symposium on Foundations of Computer Science* 124-34
[16] Gottesman D, Chuang I 2001 *Technical Report* arXiv:quant-ph/0105032
[17] Zhang J 2012 *Journal of Networks* **7**(11) 1803-10
[18] Nikolopoulos G 2008 *Physical Review A* **77** 032348
[19] Nikolopoulos G, Ioannou L 2009 *Physical Review A* **79** 042327
[20] Ioannou L, Mosca M 2009 arXiv:quant-ph/0903.5156
[21] Ioannou L, Mosca M 2011 *Proceedings of the 6th Conference on the Theory of Quantum Computation, Communication and Cryptography* 13-27
[22] Seyfarth U, Nikolopoulos G, Alber G 2012 *Physical Review A* **85** 022342
[23] Michler M, Mattle K, Weinfurter H, Zeilinger A 1996 *Physical Review A* **53**(3) 1209-12

Authors

Xiaoyu Li

Current position, grades: associate professor at the School of Information Engineering, Zhengzhou University, China.
University studies: Graduated from the Institute of Computing Technology, Chinese Academy of Sciences in 2004.
Scientific interest: quantum information and quantum computation.
Publications: more than 20.

Dai wang

Current position, grades: graduate student at the School of Information Engineering, Zhengzhou University, China.
University studies: School of Information Engineering after graduation at the School of Software, Zhengzhou University.
Scientific interest: quantum information and quantum computation.

Dual-hop variable gain relaying in mixed multipath/shadowing fading channels

Weijun Cheng*

School of Information Engineering, Minzu University of China, No. 27 South Zhongguancun Street, Haidian District, Beijing 100081, China

Abstract

In this paper, we investigated the end-to-end performance of a dual-hop variable gain relaying system over mixed fading environment. In such environment, the wireless links of relaying system undergo different fading conditions, where one link is subject to the Nakagami-m fading, the other link is subject to the composite Nakagami-lognormal fading which is approximated by using mixture gamma fading model. Based on the cumulative distribution function of the end-to-end signal-to-noise ratio (SNR), some novel closed-form expressions of the average end-to-end SNR, the outage probability, the symbol error rate and the ergodic capacity for the dual-hop variable gain system are derived, respectively. Then, some approximate analysis and the diversity order are found based on the above new expressions in high SNR region. Finally, numerical and simulation results are shown to verify the accuracy of the theoretical analysis.

Keywords: dual-hop relaying, mixed fading channels, mixture Gamma distribution, performance analysis

1 Introduction

Cooperative relaying transmission has emerged as a promising technique for extending coverage, enhancing connectivity, and saving transmitter power in wireless communications networks. In a cooperative relaying network, a source communicates with the destination via one or several intermediate terminals called relays. In the past few years, a great deal of attention has been devoted to study the performance analysis of cooperative relaying systems in term of outage probability (OP), average bit/symbol error rate (ABER/ASER) and ergodic capacity over different fading environments, such as Rayleigh, Nakagami-m, Rician, Weibull, Lognormal, Generalized Gamma and so on, e.g., [1-5] and the references therein. The common characteristic of these works is that they consider only multipath or shadowing fading channels.

In a practical scenario, relaying nodes (R) are usually located in different geographical locations and at different distances with respect to the source node (S) and the destination node (D). The signal in one link may be in line of sight (LOS) situation and other links may be in NLOS situation, even in shadowing or composite multipath/shadowing situation. In the published literature, such scenario has been regarded as asymmetric or mixed fading models [6, 7]. On the contrary, the channel situation in [1-5] is regarded as symmetric fading models in which all the single-hop links experience the same fading conditions. So far, most previous works have considered the latter, only a few works have evaluated the former [5-16]. Recently, there is an increasing research interest on the former.

In [8], we studied the performance of two-hop and multihop relaying links in random (i.e., mixed Rayleigh/ Nakagami-m) fading channels. The authors in [6] and [7] first studied the end-to-end performance of dual-hop AF relaying with both channel state information (CSI) based and fixed gain over mixed Rayleigh and Rician fading channels, respectively. After that, more cooperative relaying models are studied in mixed Rayleigh and Rician fading channels, for example, the dual-hop decode-and-forward (DF) cooperative model in [9], the dual-hop AF cooperative model in [10], the repetition-based and opportunistic amplify-and-forward (AF) model in [11], the two-hop AF networks with beamforming in [12], and so on. In [13], the authors analysed the performance of dual-hop AF relaying in mixed Nakagami-m and Rician fading channels. In [14,15], the authors studied the performance of a dual-hop DF system and an AF cooperative system under mixed Rayleigh and generalized Gamma fading channels, respectively.

Despite these recent studies related to the analysis of AF or DF relaying over asymmetric fading channels, their fading condition is limited to the mixture of various multipath fading, i.e., Rayleigh/Rician, Rician/Nakagami-m and Rayleigh/generalized Gamma fading. The performance of the cooperative networks has not as yet been widely studied under mixed multipath/shadowing fading conditions except [16, 17]. The authors in [16,17] investigated the performance of the dual-hop fixed gain relaying system over Nakagami/Generalized-K (KG) fading channels. Nevertheless, since the probability density function (PDF) of average signal-to-noise ratio (SNR) over KG fading channel includes modified Bessel functions, the outage probability and average SER in [16]

*****Corresponding author** e-mail: weijuncheng@163.com

include Meijer's G functions and some infinite-series representations. The end-to-end Moment generating function (MGF) in [17] includes Lommel function. Some expressions still keep complicated and intractable.

Recently, the authors in [18] developed an alternative approach to approximate the Nakagami-lognormal (NL) distribution by using the mixture gamma (MG) distribution. This distribution avoids the above-mentioned problems, and some exact results obtained are possible by adjusting some parameters. To the best of our knowledge, there are few papers in performance analysis of cooperative system over mixed fading channels by using MG fading model.

In this paper, we consider an asymmetric scenario of a dual-hop AF relaying system in a wireless propagation environment where multipath fading, shadowing and the propagation path loss occur simultaneously. The S−R (first-hop) and the R−D (second-hop) links experience Nakagami-m or NL fading, where the NL fading model is approximated by using MG fading model. The primary contributions of this paper are as follows: Firstly, some exact closed-form expressions of the average end-to-end SNR, the OP, the ASER and the ergodic capacity for the dual-hop system over mixed Nakagami/MG fading channels are derived, respectively. And then, some approximate analysis of the above performance is discussed and the diversity order is obtained in high SNR region. Finally, the numerical and simulation results are given to show the accuracy of the theoretical analysis.

2 System and channel model

We consider a classical wireless dual-hop variable gain relaying system consisting of S, D and R. The whole transmission is divided into two phase. In the first phase, S only transmits its signals to R, and in the second phase, R amplifies the received signal by a gain factor β and then forwards their amplified versions to D. Thus, the instantaneous end-to-end SNR, γ_{SRD}, at the destination can be expressed as in [1, 2]:

$$\gamma_{SRD} = \frac{(P_1|h_1|^2/N_0)(P_2|h_2|^2/N_0)}{[(P_2|h_2|^2/N_0) + (1/N_0\beta^2)]}, \quad (1)$$

where P_1 and P_2 are the transmitted power at S and R respectively, $|h_i|$ is the fading amplitude of the ith-hop link, $i \in \{1,2\}$, N_0 is the power of the additive white Gaussian noise component. If β is selected according to the instantaneous CSI assisted relay gain, then γ_{SRD} in Equation (1) can be re-expressed as in [1]:

$$\gamma_{SRD} = \frac{\gamma_1\gamma_2}{(\gamma_1 + \gamma_2 + c)}, \quad (2)$$

where $\gamma_i = \rho_i|h_i|^2$ is the instantaneous SNR of the ith-hop link, $\rho_i = P_i/N_0$ denotes the un-faded SNR. In addition, exact γ_{SRD} is given by substituting $c=1$ when $\beta_2 = 1/(P_1|h_1|^2+N_0)$,

and well approximated at medium and high SNR by substituting $c=0$ when $\beta_2 = 1/(P_1|h_1|^2)$).

Note that due to the symmetry of γ_{SRD} in Equation (2) with respect to γ_1 and γ_2, the statistics of γ_{SRD} will be identical despite that the ith-hop link is subject to Nakagami-m or NL fading. In a practical scenario, it is possible that a mobile station at the edge of the cell communicates with the base station via one or more mobile/fixed relaying stations over mixed multipath/shadowing situation.

If the ith-hop link experiences Nakagami-m fading, γ_i is a Gamma distributed variable with the PDF given in [19]:

$$f_{\gamma_i}(\gamma) = (m_i/\overline{\gamma}_i)^{m_i}\gamma^{m_i-1}\exp[-m_i\gamma/\overline{\gamma}_i]/\Gamma(m_i), \quad (3)$$

where $\overline{\gamma}_i = \rho_i\mathbf{E}[|h_i|^2] = \rho_i\Omega_i$ is the average SNR of the ith-hop link, m_i is Nakagami-m fading parameter, $\mathbf{E}(*)$ denotes the statistical expectation, $\Gamma(*)$ is the standard Gamma function. Due to capture the path-loss effect, we use the local mean power $\Omega_i = (d_0/d_i)^\varepsilon$, d_0 denotes the distance between S and D, d_i is the distance of the ith-hop link, and ε is the path-loss exponent.

The cumulative distribution function (CDF) of γ_i, defined as $F_\gamma(\gamma) = \int_0^\gamma f_\gamma(x)dx$ in [19], can be obtained as

$$F_{\gamma_i}(\gamma) = 1 - \Gamma(m_i, m_i\gamma/\overline{\gamma}_i)/\Gamma(m_i), \quad (4)$$

where $\Gamma(*;*)$ is the upper incomplete gamma function defined in [20].

When the ith-hop link experiences NL fading, γ_i is a composite Gamma-lognormal distribution variable with the PDF given by [19]:

$$f_{\gamma_i}(\gamma) = \int_0^\infty \frac{m_i^{m_i}\gamma^{m_i-1}\exp(-m_i\gamma/\rho_i y)}{\Gamma(m_i)(\rho_i y)^{m_2}}\frac{1}{\sqrt{2\pi}\lambda_i y}$$
$$\exp\left[-\frac{(\ln y - \mu_i)^2}{2\lambda_i^2}\right]dy, \quad (5)$$

where μ_i and λ_i are the mean and the standard deviation of lognormal shadowing, respectively, $\lambda_i = (\ln10/10)\sigma$, $\mu_i = \ln\Omega_i$, σ denotes the standard deviation in dB.

Since a closed-form expression of Equation (5) is not available in the published literature, the performance metrics of digital communication systems over composite NL distribution is intractable or difficult, some simple forms or approximations of Equation (5) have been given great attention recently, such as, KG distribution and MG distribution. Due to that KG distribution includes modified Bessel functions, some expressions of the performance metrics still keep mathematical complications, and further approximations have to be adopted. In order to avoid the above problems, we use MG distribution proposed by [18] to approximate the composite NL distribution in this paper. Thus, the PDF of γ_i can be expressed as:

$$f_{\gamma_i}(\gamma) = \sum_{j=1}^{N}(Ca_j/2\rho_i^{m_i})\gamma^{m_i-1}\exp(-b_j\gamma/\rho_i), \qquad (6)$$

where $a_j = 2m_i^{m_i}w_j\exp[-m_i(\sqrt{2}\lambda_i t_j + \mu_i)]/\sqrt{\pi}\Gamma(m_i)$,

$b_j = m_i\exp[-(\sqrt{2}\lambda_i t_j + \mu_i)]$, C is the normalization factor,

defined as $C = \sqrt{\pi}/\sum_{j=1}^{N}w_j$, w_j and t_j are abscissas and

weight factors for Gaussian-Hermite integration. w_j and t_j for different N values are available in [21, table (25.10)].

Based on the definition of the CDF, The CDF of γ_i over MG fading can be obtained as:

$$F_{\gamma_i}(\gamma) = 1 - \sum_{j=1}^{N}(Ca_j/2b_j^{m_j})\Gamma(m_i, b_j\gamma/\rho_i). \qquad (7)$$

3 Performance analysis

In this section, first we find the closed-form CDF of the end-to-end SNR for the dual-hop system, then derived the closed-form expressions of the average end-to-end SNR, the OP, the ASER and the ergodic capacity over mixed Nakagami-m/MG fading channels, respectively. Finally, some approximate analysis of the OP and the ASER is discussed and the diversity order is obtained in high SNR region.

3.1 CDF OF END-TO-END SNR

For the dual-hop system, assuming that the first-hop link is subject to Nakagami-m fading and the second-hop link is subject to NL fading, by using Equation (2), the CDF of γ_{SRD} can be expressed as in [6]:

$$F_{\gamma_{SRD}}(x) = \Pr(\gamma_{SRD} \le x) = \Pr[\gamma_1\gamma_2/(\gamma_1 + \gamma_2 + c) \le x], \qquad (8)$$

After applying some algebraic manipulations, Equation (8) can be rewritten as:

$$F_{\gamma_{SRD}}(x) = 1 - \int_0^{\infty}\bar{F}_{\gamma_1}[x + (x^2 + cx)/y]f_{\gamma_2}(x + y)dy, \qquad (9)$$

where $\bar{F}_{\gamma_1}(*)$ is the complementary CDF of γ_1, which is defined as $\bar{F}_{\gamma_1}(*) = 1 - F_{\gamma_1}(*)$. According to Nakagami-m fading of the first-hop link, $\bar{F}_{\gamma_1}(*)$ can be expressed by using Equation (4) as:

$$\bar{F}_{\gamma_1}(x) = \Gamma[m_1, M(x + (x^2 + cx)/y)]/\Gamma(m_1), \qquad (10)$$

where $M = m_1/\bar{\gamma}_1$.

By substituting Equations (6) and (10) into Equation (9), and with the help of the series expression of $\Gamma(*;*)$ defined in [20, eq. 8.352.2)] when m_i is an integer, and the binomial expansion defined in [20, eq. (1.111)], we can obtain the CDF of γ_{SRD} as:

$$F_{\gamma_{SRD}}(x) = 1 - \sum_{i=1}^{N}\sum_{k=0}^{m_1-1}\sum_{s=0}^{m_2-1}\sum_{j=0}^{k}\Xi(i,k,s,j)x^{k+(m_2+s-j)/2} \times$$
$$(x+c)^{(m_2-s+j)/2}\exp[-\Phi(i)x]K_\nu\left[\Theta(i)\sqrt{x^2+cx}\right], \qquad (11)$$

where $\Phi(i) = M + M_i$, $\Theta(i) = 2\sqrt{MM_i}$, $M_i = b_i/\rho_2$,

$\nu = m_2 - s - j$, $\Xi(i,k,s,j) = \dfrac{C_{m_2-1}^{s}C_k^{j}Ca_iM^{k+\nu/2}}{\left(b_i^{\nu/2}\rho_2^{(m_2+s+j)/2}k!\right)}$,

$C_j^i = j!/[(j-i)!i!]$ is the binomial coefficients, $K_\alpha(*)$ is the second kind modified Bessel function of order α defined in [20, eq.(8.407.1)].

3.2 OUTAGE PROBABILITY

The OP is an important performance measure that is commonly used to characterize a wireless communication system. It is defined as the probability that the instantaneous end-to-end SNR falls below a given threshold (γ_{th}), this is $P_{out} = \int_0^{\gamma_{th}}f_\gamma(\gamma)d\gamma$, where $f_\gamma(\gamma)$ is the PDF of the instantaneous SNR. Using (11), the OP of the dual-hop system over mixed fading channels can be obtained as:

$$P_{out-1} = F_{\gamma_{SRD}}(\gamma_{th}). \qquad (12)$$

3.3 AVERAGE END-TO-END SNR

The average end-to-end SNR is a useful performance measure serving as an excellent indicator of the overall system's fidelity. The qth moment of the end-to-end SNR can be derived by using CDF as:

$$\mu(\gamma^q) = \int_0^{\infty}\gamma^q f_{\gamma_{SRD}}(\gamma)d\gamma = q\int_0^{\infty}\gamma^{q-1}[1 - F_{\gamma_{SRD}}(\gamma)]d\gamma. \qquad (13)$$

By using Equations (11) with $c=0$, which is analytically more tractable and with the help of [20, eq. (6.621.3)], Equation (13) can be re-expressed as:

$$\mu(\gamma^q) = \sum_{i=1}^{N}\sum_{k=0}^{m_1-1}\sum_{s=0}^{m_2-1}\sum_{j=0}^{k}\Psi(i,k,s,j) \times$$
$$_2F_1[u_1+\nu, \nu+0.5; u_1+0.5; \frac{\Phi(i)-\Theta(i)}{\Phi(i)+\Theta(i)}] \qquad (14)$$

where

$$\Psi(i,k,s,j) = \frac{4^\nu\sqrt{\pi}C_{m_2-1}^{s}C_k^{j}Ca_iqM^{k+\nu}\Gamma(u_1+\nu)\Gamma(u_1-\nu)}{k!\rho_2^{m_2}\Gamma(u_1+0.5)[\Phi(i)+\Theta(i)]^{u_1+\nu}},$$

$u_1 = m_2 + k + q$, $_2F_1(a,b;c;z)$ is the hypergeometric function defined in [20, eq. (9.100)].

Therefore, the average end-to-end SNR can be obtained by setting $q=1$ in Equation (14).

3.4 AVERAGE SYMBOL ERROR RATE

The ASER is a useful measurement for investigating the performance of wireless communication systems. For several modulations with Gray bit mapped constellations, a uniform expression of the ASER can be written as equation [19]:

$$P_s = E[aQ(\sqrt{2b\gamma})] = \int_0^\infty aQ(\sqrt{2b\gamma}) f_\gamma(\gamma) d\gamma, \qquad (15)$$

where $Q(*)$ is the Gaussion Q-funtion defined by $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2) dt$, the parameters a and b change by specific modulation scheme. Equation (15) provides exact SER results for binary PSK (a=1, b=1), binary frequency shift keying (BFSK) (a=1, b=0.5) and M-ary pulse amplitude modulation (M-PAM) (a=2(M-1)/M, b=3/(M²-1)). Furthermore, Equation (15) also provides approximate SER results for other modulations such as M-PSK (a=2, b=sin2(π/M)). After integration by parts Equation (15) can be rewritten using the CDF of γ_{SRD}, as:

$$P_s = (a\sqrt{b}/2\sqrt{\pi}) \int_0^\infty x^{-1/2} \exp(-bx) F_{\gamma_{SRD}}(x) dx. \qquad (16)$$

Therefore, similar to Equation (14) one can get the analytical expression of ASER for the dual-hop system as:

$$P_{s-1} = \frac{a}{2} - \sum_{i=1}^N \sum_{k=0}^{m_1-1} \sum_{s=0}^{m_2-1} \sum_{j=0}^k \Upsilon(i,k,s,j) \times$$
$$_2F_1\left[u_2+\nu, \nu+0.5; u_2+0.5; \frac{\Phi(i)+b-\Theta(i)}{\Phi(i)+b+\Theta(i)}\right], \qquad (17)$$

where:

$$\Upsilon(i,k,s,j) = \frac{C_{m_2-1}^s C_k^j 4^{\nu-0.5} a\sqrt{b} Ca_i M^{k+\nu} \Gamma(u_2+\nu)\Gamma(u_2-\nu)}{k! \rho_2^{m_2} \Gamma(u_2+0.5)[\Phi(i)+b+\Theta(i)]^{u_2+\nu}},$$

$$u_2 = m_2 + k + 0.5.$$

3.5 ERGODIC CAPACITY

For a dual-hop variable gain system with the single relay, the ergodic capacity can be obtained as in [22]:

$$\overline{C} = \Delta\mathbf{E}\left[\ln(1+\gamma_{SRD})\right], \qquad (18)$$

where $\Delta = 1/2\ln 2$.

Since an exact closed-form expression in Equation (18) over mixed Nakagami-m/MG fading channels is not mathematically tractable by directly using a traditional approach (i.e., finding the PDF of γ_{SRD} with c=1), we thus restructure Equation (18) as in[22]:

$$\overline{C} = \Delta\{\underbrace{\mathbf{E}[\ln(1+\gamma_1)]}_{\tilde{C}_1} + \underbrace{\mathbf{E}[\ln(1+\gamma_2)]}_{\tilde{C}_2} - \underbrace{\mathbf{E}[\ln(1+\gamma_1+\gamma_2)]}_{\tilde{C}_3}\}. \quad (19)$$

Note that Equation (19) provides an interesting information. Theoretic result that states that the ergodic capacity of the dual-hop system is equal to the sum of the ergodic capacities of the source–relay link $\overline{C}_1$ and relay–destination link $\overline{C}_2$ minus the ergodic capacity $\overline{C}_3$ of the single input multiple output system, in which the source acts as a transmitter, and the relay and destination are the receivers. The advantage of Equation (19) is now clear, because the methods of finding closed-form expressions for the ergodic capacities are already available in the open literature, i.e., using the PDF of γ_i. In the following, we will now derive new closed-form expressions for the ergodic capacities of the dual-hop variable gain relay network over mixed Nakagami-m/MG fading channels by using Equation (19).

First, we find the closed-form expressions of $\overline{C}_1$ and $\overline{C}_2$. By using Equation (3) and the integral expression in [19, eq.(15B. 7)] as:

$$\Im_n(\mu) = \int_0^\infty t^{n-1} \ln(1+t)\exp(-\mu t) dt =$$
$$(n-1)!\exp(\mu)\sum_{k=1}^n \Gamma(-n+k,\mu)/\mu^k, \mu>0, n=1,2,...,$$

then the closed-form expression of $\overline{C}_1$ can be written as

$$\overline{C}_1 = \int_0^\infty \ln(1+x) f_{\gamma_1}(x) dx = M^{m_1} \Im_{m_1}(M)/\Gamma(m_1). \qquad (20)$$

Similarly, the closed-form expression of $\overline{C}_2$ can be written as:

$$\overline{C}_2 = \int_0^\infty \ln(1+x) f_{\gamma_2}(x) dx = \sum_{i=1}^N Ca_i \Im_{m_2}(M_i)/2\rho_2^{m_2}. \quad (21)$$

Then, we find the closed-form expression of $\overline{C}_3$. Here, we let $z = \gamma_1 + \gamma_2$. By using Equation (3) and (6), the PDF of variable z can be obtained as:

$$f_z(z) = \sum_{i=1}^N [Ca_i M^{m_1}/2\rho_2^{m_2}\Gamma(m_1)]\exp(-M_i z)\times$$
$$\int_0^z x^{m_1-1}(z-x)^{m_2-1}\exp\left[-x(M-M_i)\right] dx \qquad (22)$$

When $M \neq M_i$, by using the binomial expansion defined in [20, eq.(1.111)] and eq.(3.381) in [20], after applying some algebraic manipulations, Equation (22) can be rewritten as:

$$f_z(z) = \sum_{i=1}^N \sum_{j=0}^{m_2-1} \frac{Ca_i M^{m_1}(-1)^j C_{m_2-1}^j}{2\rho_2^{m_2}\Gamma(m_1)(M-M_i)^{m_1+j}} \times$$
$$z^{m_2-j-1}\exp(-M_i z)\gamma\left[m_1+j,(M-M_i)z\right] \qquad (23)$$

where $\gamma(*,*)$ is the lower incomplete gamma function defined in [20, eq.(8.350.1)].

With the help of the series expression of $\gamma(*,*)$ defined in [20, eq. 8.352.1)], and similar to the Equation (20), after applying some algebraic manipulations, the closed-form expression of $\overline{C}_3$ when $M \neq M_i$ can be written as:

$$\overline{C}_3 = \sum_{i=1}^{N} \sum_{j=0}^{m_2-1} \frac{Ca_i M^{m_1}(-1)^j C_{m_2-1}^j \Gamma(m_1+k)}{2\rho_2^{m_2}\Gamma(m_1)(M-M_i)^{m_1+j}} \times$$
$$\left(\Im_{m_2-j}(M_i) - \sum_{k=0}^{m_1+j} \frac{(M-M_i)^k}{k!} \Im_{m_2+k-j}(M) \right). \quad (24)$$

When $M = M_i$, by using equation (3.191.1) in [20], (22) can be rewritten as:

$$f_z(z) = \sum_{i=1}^{N} \frac{Ca_i M^{m_1}\Gamma(m_2)}{2\rho_2^{m_2}\Gamma(m_1+m_2)} z^{m_1+m_2-1} \exp(-M_i z). \quad (25)$$

Similarly, the closed-form expression of $\overline{C}_3$ when $M = M_i$ can be written as:

$$\overline{C}_3 = \sum_{i=1}^{N} \frac{Ca_i M^{m_1}\Gamma(m_2)}{2\rho_2^{m_2}\Gamma(m_1+m_2)} \Im_{m_1+m_2}(M_i). \quad (26)$$

Finally, by substituting Equations (20), (21) and (24) or Equation (26) into Equation (19), we can obtain the exact closed-form expression for the ergodic capacity of the dual-hop system over mixed Nakagami-m/MG fading channels.

3.6 DIVERSITY ORDER ANALYSIS

Performance results obtained for OP and SER expressions in Equations (12) and (17) do not reveal any information about the diversity order and array gain of the relaying system. Therefore, we first try to find the upper bound of the OP and SER performance in this section, and then obtain their approximate performance in the high SNR region. Recently, in order to simplify the performance analysis of Equation (2) over Nakagami-m, Weibull and KG fading, its looser upper bound is often adopted in many recent literatures as [3]

$$\gamma_{SRD} < \gamma_b = \min(\gamma_1, \gamma_2). \quad (27)$$

The physical interpretation of the upper bound SNR in (27) is that at high SNR region, the hop with the weaker SNR determinates the end-to-end system performance. This upper bound has been shown to be accurate enough at high SNR region. Based on (27), the OP of the dual-hop system can be expressed as:

$$P_{out-\gamma_b} = F_{\gamma_1}(\gamma_{th}) + F_{\gamma_2}(\gamma_{th}) - F_{\gamma_1}(\gamma_{th})F_{\gamma_2}(\gamma_{th}), \quad (28)$$

where $F_{\gamma_i}(*)$ ($i=1,2$) is the CDF of the ith-hop link, and can be found using Equations (4) and (7). Substituting them into Equation (28) the following equation is received:

$$P_{out-2} = 1 - \sum_{i=1}^{N} \frac{Ca_i}{2b_i^{m_2}\Gamma(m_1)} \Gamma(m_1, M\gamma_{th})\Gamma(m_2, M_i\gamma_{th}). \quad (29)$$

Although Equation (12) is valid and accurate and Equation (29) is its upper bound, they are too complicated to provide a clear insight about the diversity order and

array gain of the system. Thus, in what follows, we derive the approximate OP expression at high SNR to reveal them for the dual-hop relaying system.

Since the values of $F_{\gamma_i}(*)$ in Equation (17) range between 0 and 1, the product of these two CDFs will be much less than their addition when $\rho_i \to \infty$. Hence, by neglecting the product term in Equation (28) and using the series expression of $\Gamma(*;*)$ defined in [20, eq. (8.354.2)] in high SNR region, we can derive an approximating OP when $\rho_1 = \rho_2 = \rho \to \infty$, as:

$$P_{out-3} \approx F_{\gamma_1}(\gamma_{th}) + F_{\gamma_2}(\gamma_{th}) \approx A(\gamma_{th}/\rho)^t + O[(\gamma_{th}/\rho)^{t+1}], (30)$$

where $O[|x|^{t+1}]$ represents the terms of order higher than t, $t = \min(m_1, m_2)$,

$$A = \begin{cases} \sum_{i=1}^{N} \dfrac{Ca_i}{2m_2} & m_1 > m_2 \\[2ex] \dfrac{(m/\Omega_1)^m}{m\Gamma(m)} + \sum_{i=1}^{N} \dfrac{Ca_i}{2m} & m_1 = m_2 = m \\[2ex] \dfrac{(m_1/\Omega_1)^{m_1}}{m_1\Gamma(m_1)} & m_1 < m_2 \end{cases}$$

Similarly, by setting $\gamma_{th} = x$ in Equation (29) and substituting Equation (29) into Equation (16), the upper bound expression of SER can be obtained as:

$$P_{s-2} = \frac{a}{2} - \sum_{i=1}^{N} \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \frac{a\sqrt{b}Ca_i\Gamma(m_2)\Gamma(0.5+j+k)M^j}{4\sqrt{\pi}\,j!k!\rho_2^k(b_i)^{m_2-k}\xi(i)^{0.5+j+k}}. \quad (31)$$

To find the asymptotic analysis of the SER, P_s can be expressed as:

$$P_{s-3} = (G_a\rho)^{-G_d}, \quad (32)$$

where G_a and G_d are the array gain and diversity order, respectively. By substituting Equation (30) into Equation (16), the diversity order of the dual-hop system can be given as $G_d = \min(m_1, m_2)$ when $\rho_1 = \rho_2 = \rho \to \infty$, the array gain can be expressed as:

$$G_a = \begin{cases} \left(\dfrac{a\Gamma(m_1+1)(m_1/\Omega_1)^{m_1}}{2m_1\sqrt{\pi}\Gamma(m_1)b^{m_1}} \right)^{-1/m_1} & m_1 < m_2 \\[3ex] \left(\dfrac{a\Gamma(m+1)(m/\Omega_1)^m}{2m\sqrt{\pi}\Gamma(m)b^m} + \sum_{i=1}^{N} \dfrac{aCa_i\Gamma(m+0.5)}{4m\sqrt{\pi}b^m} \right)^{-1/m} & m_1 = m_2 = m \\[3ex] \left(\sum_{i=1}^{N} \dfrac{aCa_i\Gamma(m_2+0.5)}{4m_2\sqrt{\pi}b^{m_2}} \right)^{-1/m_2} & m_1 > m_2 \end{cases} \quad (33)$$

For Equations (30) and (32), we can obtain the diversity order of the dual-hop system is $\min(m_1, m_2)$, which implies that the weaker hop dominates the system performance.

4 Numerical and simulation results

In this section, we present some numerical and simulation results to evaluate the performance of the dual-hop system in mixed fading channels, where the S-R link is subject to the Nakagami-m fading, R-D link is subject to the MG fading, vice versa.
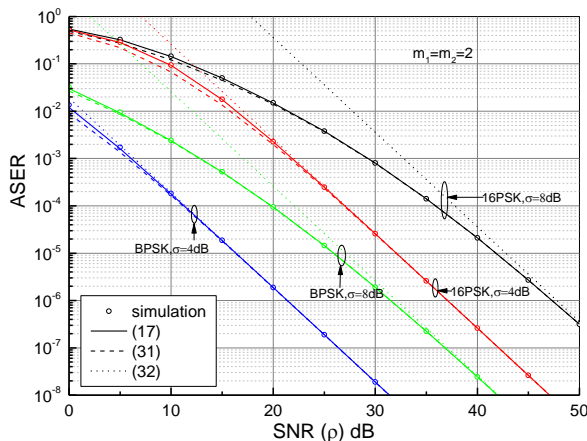
FIGURE 1 ASER of BPSK and 16PSK for the dual-hop system versus the unfaded SNR (ρ)
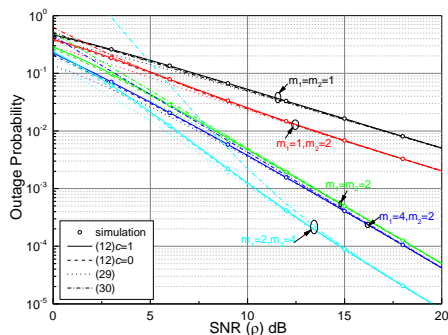
FIGURE 2 Outage probability for the dual-hop system versus the unfaded SNR (ρ)

We show the effect of the relay location on the ABER of BPSK for the dual-hop system in Figure 3. In this section, the asymmetric network geometry is examined where R is moved on a straight line between S and D, d_1 denotes the distance between S and R. Each hop has different fading parameters, $\rho_1 = \rho_2 = 10$dB, $N = 10$ for MG distribution. It can be seen from Figure 3 that the optimum performance of the dual-hop system moves with the channel parameters of the weaker hop. When R is closer to S, the system performance is determined by the channel condition of the second hop, for example, if the fading parameter increases ($m_2 = 2 \rightarrow m_2 = 4$), the performance is improved and the location of the optimum performance moves toward S, and if the shadow deviation increases

Figure 1 illustrates the ASER of BPSK and 16PSK of the dual-hop variable gain system. The analytical results in Equations (17), (31), (32) and the simulation results ($c=1$) are given, respectively. In this case, a symmetric network geometry is assumed, this is, $d_0 = 1$, $d_1 = d_2 = 0.5$, $\varepsilon = 4$, $\rho_1 = \rho_2 = \rho$. Each hop has the same fading parameters ($m_1 = m_2 = 2$), $N = 10$ for MG distribution. It can be seen from Figure 1 that the analytical results in Equation (17) has almost the same as the ones in Equation (31) in high SNR region, only a small gap in low SNR region. At the same time, the analytical results of Equation (17) coincide perfectly with the simulation results. The slops of approximate performance in Equation (32) show agreement with Equation (17) in high SNR region. As expected, the system performance is degraded when the shadow deviation increases, and the performance of BPSK outperforms that of 16PSK at the same channel conditions.

The analysis and simulation results of OP are shown in Figure 2, where $d_1 = d_2 = 0.5$, $\sigma = 4$dB, $\rho_1 = \rho_2 = \rho$ and $N = 10$ for MG distribution. The analytical results in (12), (29), (30) and the simulation results are also given, respectively. As expected, these results are similar as that in Figure 1. Moreover, for discussing their diversity orders, it can be seen the diversity order of the dual-hop is determined by the minimum value between m_1 and m_2. It can be also seen that the fading parameter value of the second-hop increases, the array gain of the dual-hop system is improved. This is due to the fact that the system performance is determined by the weaker hop (i.e., the second-hop, $\sigma = 4$dB).
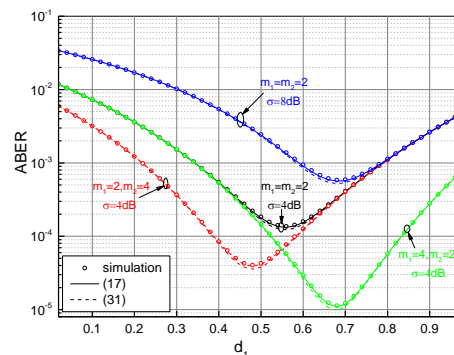
FIGURE 3 ABER of BPSK for the dual-hop system versus d_1

($\sigma = 4$dB$\rightarrow \sigma = 8$dB), the performance is degraded. When R is closer to D, the system performance is determined by the channel condition of the first hop. If the fading parameter increases ($m_1 = 2 \rightarrow m_1 = 4$), the performance is improved and the location of the optimum performance moves toward D. It can also be explained that they show the same performance when R is closer to D if only the channel conditions of the second-hop change. These results are helpful to the selection of relaying nodes in cooperative networks. Moreover, we also show the comparisons among Equations (17), (31) and the simulation results. It is clear that the difference between Equations (17) and (31) is small except the nearby region of the optimum performance. At

the same time, the simulation results show agreement with Equation (17).
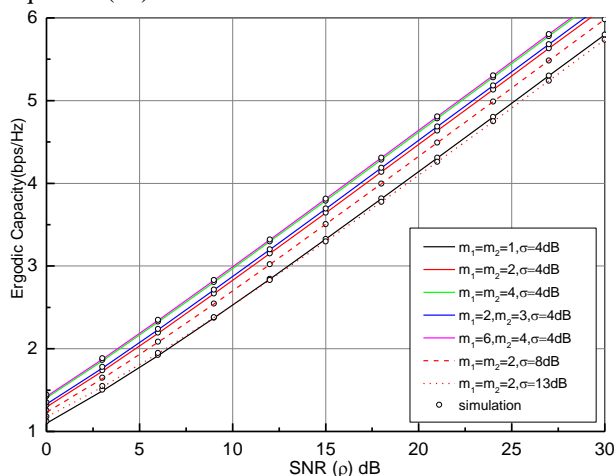
FIGURE 4 Ergodic capacity for the dual-hop system versus the unfaded SNR (ρ) under different fading parameters

5 Conclusion

In this paper, we investigated the end-to-end performance of a dual-hop AF relaying system over mixed multipath/ shadowing fading environment, where the composite NL distribution is approximated by using mixture gamma distribution. Based on the CDF of the end-to-end SNR, some novel closed-form expressions of the average end-to-end SNR, the OP, the ASER and the ergodic capacity for the dual-hop AF system are derived, respectively. And then, some approximate analysis and the diversity order are found in high SNR region. Finally, we showed numerical and simulation results to verify the accuracy of the analytical results, and discussed the effect of the location of relaying node on the performance of the dual-hop system. These results are helpful to the selection of relaying nodes in cooperative networks. Furthermore, these works in this paper can be helpful to analyse the performance of cooperative relaying systems with co-channel interference over composite fading channels by using MG fading model in the future.

References

[1] Laneman J N, Tse D N C, Wornell G W 2004 *IEEE Transactions on Information Theory* **50**(12) 3062-80
[2] Hasna M O, Alouini M S 2003 *IEEE Transactions on Wireless Communications* **2**(6) 1126-31
[3] Ikki S, Ahmed M H 2007 *IEEE Communications Letters* **11**(4) 334-6.
[4] Majid S, Murat U 2008 *IEEE Transactions on Wireless Communications* **7**(5) 1963-72
[5] Peppas K P, Mansour A, Tombras G S 2010 Dual-hop transmissions with fixed-gain relays over Generalized-Gamma fading channels *Journal of Telecommunications* **1**(1) 87-93
[6] Suraweera H A, Louie R H Y, Li Y H, Karagiannidis G K 2009 *IEEE Communications Letters* **13**(4) 227-9
[7] Suraweera H A, Karagiannidis G K, Smith P J 2009 *IEEE Transactions on Wireless Communications* **8**(6) 2783-8
[8] Cheng W J, Zhu B C, Hu J D 2005 Analysis on the performance of multi-hop wireless transmissions over random fading channels *Journal of Beijing University of Posts and Telecom-munications* **5**(A) 5-9 *(in Chinese)*
[9] Yang C Q, Wang W B, Zhao S A, Peng M G 2011 Opportunistic decode-and-forward cooperation in mixed Rayleigh and Rician fading channels *ETRI Journal* **33**(2) 287-90
[10] Ouyang J, Lin M, Zhang Y 2012 Performance analysis of cooperative relay networks over asymmetric fading channels *Electronics Letters* **48**(21) 1370-1
[11] Nasser Y, Majhi S, Herald J-F 2010 Outage probability analysis of cooperative communications over asymmetric fading channel *Communications and Networking Jun Peng (Ed.)* ISBN: 978-953-307-114-5 InTech DOI: 10.5772/10165
[12] Chen S H, Liu F, Zhang X, Han Y N, Yang D C 2010 On the performance of two-hop amplify and forward relay networks with beamforming over Rayleigh-Rician fading channels *Proceedings of IEEE 72nd VTC (Fall)* Ottawa 1-4
[13] Xu W, Zhang J, Zhang P 2010 Performance analysis of dual-hop amplify-and-forward relay system in mixed Nakagami-m and Rician fading channels *Electronics Letters* **46**(17) 1231-2
[14] Kapucu N, Bilim M, Develi I 2013 SER performance of amplify and forward cooperative diversity over asymmetric fading channels *Wireless Personal Communications* **73**(3) 1117-27
[15] Kapucu N, Bilim M, Develi I 2013 Outage probability analysis of dual-hop decode-and-forward relaying over mixed rayleigh and generalized gamma fading channels *Wireless Personal Communications* **71**(2) 947-54
[16] Trohui I, Affes S, Stéphenne A 2010 On the performance of dual-hop fixed gain relaying systems over composite multipath/ shadowing channels *Proceedings of IEEE 72nd VTC (Fall)* Ottawa 1-5
[17] Waqar O, Imran M A, Dianati M 2013 On the error analysis of fixed-gain relay networks over composite multipath/shadowing channels *Proceedings of IEEE 77th VTC (Spring)* Dresden 1-5
[18] Atapattu S, Tellambura C, Jiang H 2010 Representation of composite fading and shadowing distributions by using mixtures of gamma distributions *Proceedings of IEEE Wireless Communications and Networking Conference* Sydney 1-5
[19] Simon M K, Alouini M S 2005 Digital Communication over Fading Channels 2nd edition *New York Wiley*
[20] Gradshteyn I S, Ryzhik I M 2000 Table of Integrals, Series, and Products, 6th Edition *New York Academic*
[21] Abramowitz M, Stegun I A 1965 Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables *Dover Publications*
[22] Waqar O, McLernon D, Ghogho M 2010 *IEEE Transactions on Vehicular Technology* **59**(8) 4181-7

Author

Weijun Cheng, born in November, 1973, Shandong, China

Current position: PhD of communications Engineering, Associate professor.
University studies: MSc degree in Electronics and Control Engineering at China University of Mining and Technology, Beijing, China, in 1998. PhD degree in Telecommunications Engineering at Beijing University of Posts and Telecommunications, Beijing, China, in 2004.
Scientific interest: wireless cooperative communication theory and radio resource management.
Publications: 35 papers, 2 books.
Experience: postdoctoral research fellow from 2005 to 2007 in Electronics Engineering Department from Peking University, Beijing, China.

Scalable authentication protocol in RFID-based systems

Xuping Ren[1*], Haiping Zhang[1], Yunfa Li[2], Xindong You[2]

[1]*School of information engineering, HangZhou Dianzi University, Hangzhou, China*

[2]*Institute of Software and Intelligent Technology, HangZhou Dianzi University, Hangzhou, China*

Abstract

Because of vast highly sensitive business information within RFID system, there is an urgent need for an effective and secure protocol to ensure the interests of various stakeholders. In this paper, we propose a scalable authentication protocol to provide classification protection. GNY logic formal approach is used to verify the design correctness of the protocol. The performance is evaluated and compared with other related protocols in three aspects: storage, computation requirement and communication overload. The analysis shows that the proposed protocol need less computation requirement and memory with acceptable communication overload. The conclusion indicates that the protocol is reliable and more scalable in RFID-based sensor systems.

Keywords: RFID security, authentication protocol, scalable

1 Introduction

Radio frequency identification (RFID) is an automatic identification technology. The tag communicates with a reader via wireless channels where neither visual nor physical contact is needed. RFID systems, thanks to their low cost and their convenience in identifying an object, have found many applications in manufacturing, supply chain management (SCM), parking garage management, and inventory control [1]. The wireless communication is more vulnerable to malicious adversaries. RFID system pays more attention to the tagged item's information because tags usually contain some sensitive or personal data. It is critical that the tag data can access only by authorized readers.

RFID is especially useful for SCM, whose goal is to manage all the steps involved in product manufacturing, distributing, and retailing, in order to minimize the operating expenses by managing the stock [2]. There are various stakeholders (e.g. manufacturer, material supplier, carrier and retailer) in SCM. Each stakeholder is permitted to access the authorized tag data, whereas any irrelevant sensitive data of other groups is not disclosed [3]. The RFID-based SCM systems are faced with two threats: internal attacks and external attacks. Both attacks may lead to security threats and privacy disclosure. We refer the internal attacks to internal legal entities, who may impersonate as other legal entities to do authority-exceeding violation. For example, a manufacturer's reader personates retail's reader to access a tag. We refer the external attacks to external illegal entities (such as adversaries or business competitors), who may do spoofing attack, replay attack, Denial of server attack and so on.

Many schemes have been proposed to address the potential security and privacy problems in RFID systems. These schemes can be categorized in three main types (traditional protocols, ownership transfer protocols and classification protection protocols) for RFID-based SCM systems.

Traditional protocols [4-6] use many methods to achieve the security goals. Those protocols may not be efficient or robust enough owing to various security vulnerabilities [7]. Most protocols have been designed without strict formal proof, which may detect design flaws and security vulnerabilities. Some protocols [8, 9] lack the idea of classified security protection for an overall management. Other schemes [10, 11] focus on the external illegal attacks, but ignore the forgery attacks from the internal legal entities. So such protocols are difficult to directly apply to the RFID-based SCM system.

The second type is ownership transfer protocols. Many ownership transfer protocols are proposed to fulfill security requirements of those stakeholders. The security requirements include the secure ownership, exclusive ownership and secure ownership transfer [12]. Some transfer protocols [13] adopt the asymmetric key authentication. Some RFID ownership transfer protocols are based on the symmetric key authentication schemes. According to the current ownership of the tag, a user can be a previous owner, current owner or new owner of the tag [14]. Such protocols have classified protection of ideas. However, there are many situations that are not the owner transfers in supply chain management. For instance, carrier is not the owner of anything, is only responsible for the goods delivered. In order to facilitate the management, it needs to gather some information from the tags, rather than the overall information.

---

[*]*Corresponding author* e-mail: renxp@hdu.edu.cn

Fore-mentioned schemes allow all the authorized readers can access all entire identifiers of all legal tags. Based on the previous analysis, it is essential for authenticated entities to access the specified field areas of a tag identifier (TID) [3].

To solve this problem, H. Ning et al. proposed a distributed key array authentication protocol (KAAP) [3] for RFID-based sensor systems to realize classified security protection and resist attacks. The authors declared the protocol is reliable and scalable in advanced RFID-based sensor systems. But the reader groups are predetermined. Meanwhile those grouping information need to write into tags. These features limit the protocol scalability.

Md Monzur Morshed et al. proposed three secure ubiquitous authentication protocols SUAP1, SUAP2 and SUAP3 for RFID systems. The final version of SUAP1 is improved over the preliminary version with privacy and security enhancement. SUAP2 and SUAP3 are the extension of SUAP1 and work in a large group-based system where RFID tags are divided into several groups. SUAP3 is proposed to provide a larger range of privacy and security protections for low storage and computations. It aimed at ubiquitous computing environment. But it is not suitable to be used directly on the common RFID system.

The aim of this paper is to design a classification protection and better scalable protocol for RFID-based SCM system. All readers and tags are divided into several groups. One reader can change its role successfully without tags' intervention. All the information about groups is stored in the back-end database. The main contributions of our work are as follows:

1) One-way hash function is adopted to protect PID_{Tj} and PID_{Rj} to realize no reversibility without revealing any sensitive data.

2) The authentication key array K_{MN} and the role array K_S are used to realize the classified security protection and resist internal and external attack.

3) Pseudorandom identifiers are transmitted instead of the real identifiers.

4) Access lists (L_R, L_T,) store all the pseudorandom identifiers and they are used to retrieve a certain reader or a certain tag for quick search. The hash values of the pseudorandom identifiers use to index certain reader or tag.

5) Mutual authentication procedure is performed to realize access control.

The organization of the paper is as follows. In section II, related RFID protocols are reviewed and analysed. Section III describes the proposed protocol. In next section, formal analysis of the protocol with GNY logic has been given. Performance analysis has been discussed in section V. Finally, section VI draws a conclusion.

2 Related works

In this section, two prominent protocols KAPP [3] and SUAP3 [10] will be discussed in more detail as they are more related to the proposed work. Two protocols both work in large group-based system where RFID tags are divided into several groups.

2.1 KAPP SCHEME

KAPP scheme uses distributed key array to realize classified security protection and resist both external and internal attacks, uses pseudorandom identifiers to resist the forward traceability, and adopts access lists to conserve memory and improve scalability. The notations and symbols used in KAPP operation are as follows in Table 1 [3]:

TABLE 1 The notations and symbols used in KAPP operation

Notation	Description
G_{Rm}, G_{Tn}	the mth reader group and the nth tag group in the RFID based sensor system, (m =1,2,…,M; n =1,2,…,N).
R_i	the ith reader who belongs to G_{Rm}, (i =1,2,…,I; I ≥ M).
T_j	the jth tag who belongs to G_{Tn}, (j =1,2,…,J; J ≥N).
PID_{R_i}, PID_{T_j}	the pseudorandom identifiers of R_i and T_j, which have special flags to mark R_i and T_j .
L_R	the access list for tags to retrieve a certain reader.
L_T	the access list for the database to retrieve a certain tag
r_{R_i}, r_{T_j}	the general formats of random numbers for R_i and T_j.
$r_{R_{igen}}$, $r_{T_{jgen}}$	the random numbers generated by R_i and T_j in one session.
r_{Ricom}, r_{Tjcom}	the random numbers computed by R_i and T_j in one session.
k_u	the shared key is pre shared between legal readers and tags, and it is a secure value without being revealed to the third entity.
$k_{m,n}$	the authentication key owned R_i and T_j, which is assigned to G_{Rm} and $G_{Tn.}$
‖	concatenate operator.
$\rightarrow$	transition operator.
$\{.\}_k$	encryption with key k.
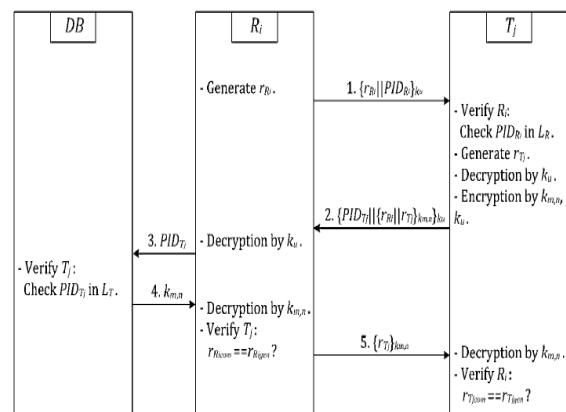
FIGURE 1 KAPP

The protocol is shown in Figure 1. All legal readers and tags are divided into different groups in the protocol. However, the reader groups are predetermined. Meanwhile those grouping information need to write into tags. These features limit the protocol scalability.

2.2 SUAP3 SCHEMES

Md Monzur Morshed et al. proposed three secure ubiquitous authentication protocols SUAP1, SUAP2 and SUAP3 for RFID systems. These protocols are low-cost and secured based on challenge-response method using a one-way hash function, hash address as a search index. The proposed protocols combine the features of the hash address and hash function of the LCAP protocol and the ubiquitous property of OHLCAP protocol. The detail about SUAP1 and SUAP2 can be found in [10]. Here we focus on SUAP3.

The notations and symbols used in SUAP3 are as follows Table 2 [10].

TABLE 2 The notations and symbols used in SUAP3

Notation	Description
ID	tag identifier
GID	group identifier
Had	hash address h(ID)
h	a one way hash function, h: $\{0,1\}^* \rightarrow \{0,1\}^l$
r_1, r_2	random number in $\{0,1\}^l$
l	the length of an identifier
$\oplus$	XOR operator
$\|$	concatenation operator
$\leftarrow$	assignment operator
h_L	the left half of $h(ID\|r_1\|r_2\|GID)$
h_R	the right half of $h(ID\|r_1\|r_2\|GID)$

SUAP3 is shown in Figure 2. SUAP3 only requires two one-way hash function operations and avoids large number of hash computations in the database. The tag search time in the database is reduced by using the hash value as the address of the corresponding tag. The authors declare that SUAP3 ensures privacy and security protections from all the identified threats. The analysis shows that the storage requirements in SUAP3 are also less than other related protocols. The comparison shows that the proposed protocol SUAP3 is both secure and efficient than other related schemes, and has practical advantages over them because it is simple and provides a larger range of privacy and security protections for low storage and computations.

However, SUAP3 is not suitable to be used directly on the common RFID system. Furthermore, whether the reader who launches a query is legal or not, the tag must give a reply. This feature will cause the occurrence of denial of service (DOS): the tag may have no time to answer the inquiry from a legitimate reader.
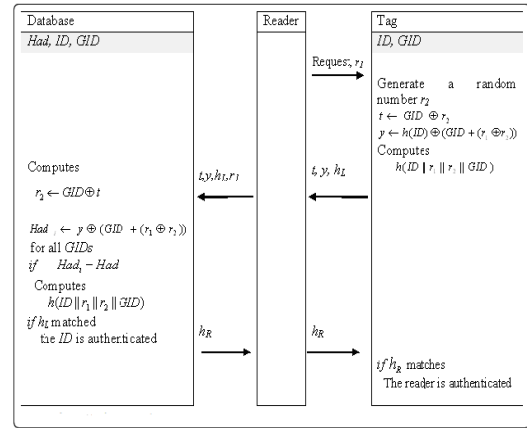
FIGURE 2 SUAP3

3 Protocol descriptions

Table 3 shows the notations applied in the protocol.

TABLE 3 Notation

Notation	Description
R	The reader in the RFID system
T	The tag in the RFID system
DB	The database in the RFID system
L_R	the access list for tags to retrieve a certain reader
L_T	the access list for the database to retrieve a certain tag
$r_R, r_T,$	the random numbers generated by R, T
$PID_R,$ PID_T	The pseudonym of R, T
h( )	A one-way hash function
[ ]	The rounding operation
$\oplus$	XOR bitwise logic operator
$\|$	Concatenate operator
K_S	the role array which stores in DB
k_{ij}	the authentication key owned R_i and T_j, which is assigned to G_{Rm} and G_{Tn} $k_{ij}=a_1,a_2,...a_s$. $a_x \in \{0,1\}$, $x=1,2,...s$. s is the number of TID fields. $a_x=1$ represents R_i is authorized to access the xth filed of T_j. Otherwise, R_i can not access the xth filed
K_{MN}	the key array which stores all the authentication keys in DB

3.1 INITIALIZATION

The system set-up of the new protocol is as follows:

Tag: each tag contains the following fields L_R, PID_{Tj} and can perform one-way hash function.

Reader: each reader contains PID_{Ri}.

Database: the back-end database contains L_T, L_R, the role array K_S and the key array K_{MN}.

3.2 AUTHENTICATION PROCESS

The protocol is summarized in Figure 3. The authentication process includes five phases: challenge messages; response messages; forward messages; authenticate the tag and authenticate the reader. Figure 3 describes the protocol in detail according to the sequence of message exchanges.
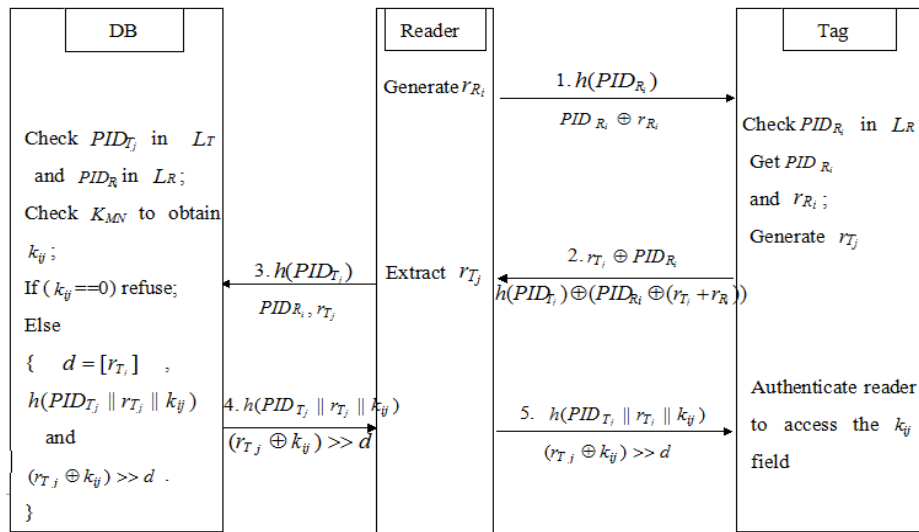
FIGURE 3 The proposed protocol

The proposed protocol based on a key array and a role array adopts lightweight mechanisms to realize security, efficiency and reliability for low-cost and large group-based RFID system. The main approaches include:

1) Mutual authentication procedure is performed to realize access control. The reader needs to be verified by the tag and DB. DB authenticates T_j and R_j, then transmits messages to R_j before the tag indeed authenticates the reader. If and only if both authentications succeed, communication between R_j and T_j is secure and will continue.

2) Pseudorandom identifiers are transmitted instead of the real identifiers. In each session, T_j and R_j generate their random numbers r_{Tj} and r_{Ri}, which are to ensure dynamic refreshment. If a new query arrives with the same data within certain time, it will be neglected. This will help the system resist the replaying or jamming attacks.

3) The role array K_S and the authentication key array K_{MN} are used to realize the classified security protection and resist attacks. The role array K_S describes the rights of various types of readers. Because any reader may change its role during the system operating, the contents of K_{MN} may be changed, that is the value of the authentication key k_{ij} owned R_j and T_j can be varied. The tag needs to do nothing in the procedure. This enables the system to be more scalable.

4) One-way hash function is adopted to protect PID_{Tj} and PID_{Rj} to realize no reversibility without revealing any sensitive data.

5) Access lists (L_R, L_T) store all the pseudorandom identifiers and they are used to retrieve a certain reader or a certain tag for quick search. Each tag maintains L_R and DB maintains both. The hash values of the pseudorandom identifiers use to index a certain reader or tag. The access lists effectively reduce the time complexity of search operation and enable better scalability for dynamic systems.

In general, the proposed protocol is based on a dynamic key array and its TID is never exposed in plain form. It is meaningful for ranking diversified authorities to realize classified security protection. In next section, GNY Logic [16] is applied to analyse the design correctness of the proposed protocol.

4 Formal analysis of the protocol with GNY logic

Most authentication protocols have been designed and demonstrated in informal ways. Design flaws and security errors may be ignored by informal analysis [11]. With the formal method, a protocol can be demonstrated to reasonably achieve its goals using logical postulates [3].

We do the GNY formal logic analysis like [3], and the analysis involves four steps:

1) formalization of the protocol messages;
2) declaration of initial assumptions;
3) definition of anticipant goals;
4) verification by logical rules and formulae.

4.1 FORMALIZATION OF MESSAGES

We express each exchanged message as a logical formula and formalization of the messages in the language of GNY Logic. For the sake of clarity, we use the same statements as [11,15]. Table 4 shows those statements.

TABLE 4 Basic statement

Notation	Description
$S \triangleleft X$	S receives a message containing X, S can read and repeat X
$S \triangleleft * X$	S receives X, X is a not-originated-here formula
$S \ni X$	S possesses, or is capable of possessing X
$S \mid\sim X$	S once conveyed X
$S \models X$	S believes, or would be entitled to believe, that statement X holds
$S \models \varphi X$	S believes, or is entitled to believe that X is recognizable
$S \models \# X$	S believes, or is entitled to believe that X is fresh
$S \models S \xleftrightarrow{v} X$	S believes, or is entitled to believe, that V is a suitable secret for S and X
$\{X, Y\}$	Concatenation

According to the authentication phase, the formalized messages are as follows:

M1 $(R_i \rightarrow T_j): h(PID_{R_i}), (PID_{R_i} \oplus r_{R_i})$;

M2 $(T_j \rightarrow R_i): r_{Tj} \oplus PID_{R_i}, \mathrm{h}(PID_{T_j}) \oplus (PID_{R_i} \oplus (r_{T_j} + r_{R_i}))$;

M3 $(R_i \rightarrow DB): h(PID_{T_j}), r_{T_j}, PID_{R_i}$;

M4 $(DB \rightarrow R_i): h(PID_{T_j} \| r_{T_j} \| k_{ij}), (r_{T_j} \oplus k_{ij}) >> d$;

M5 $(R_i \rightarrow T_j): h(PID_{T_j} \| r_{T_j} \| k_{ij}), (r_{T_j} \oplus k_{ij}) >> d$.

4.2 INITIAL ASSUMPTIONS

In order to specify the initial possessions and abilities of each participant, we assume the following statements:

(A1) $T \ni r_T$;

(A2) $T \mid\equiv \# k_{ij}$;

(A3) $T \ni PID_T, T \mid\equiv T \xleftarrow{PID_T} DB$

(A4) $T \mid\equiv \# PID_R, T \mid\equiv T \xleftarrow{PID_R} R$;

(A5) $R \ni r_R$;

(A6) $R \ni PID_R, R \mid\equiv \# PID_R, R \mid\equiv R \xleftarrow{PID_R} T$;

(A7) $R \mid\equiv DB \mid\Rightarrow (DB \mid\equiv *)$;

(A8) $DB \ni k_{ij}, DB \mid\equiv \# k_{ij}$;

(A9) $DB \ni PID_T, DB \mid\equiv \# PID_T, DB \mid\equiv DB \xleftarrow{PID_T} T$;

(A10) $DB \ni PID_R, DB \mid\equiv \# PID_R, DB \mid\equiv DB \xleftarrow{PID_R} R$.

Those statements show that each participator possesses its random number and the pseudorandom identifier. Each tag believes or is entitled to believe k_{ij} and PID_R are fresh. The database DB possess k_{ij}, PID_R and PID_T, and it believes that they are fresh. The communication channel between R and DB is considered to be secure; so R believes that DB has jurisdiction over all his beliefs.

4.3 ANTICIPANT GOALS

The objectives of the protocol are to mutually authenticate between R and T, and assure freshness of data among R, T and DB. The anticipant goal can be obtained as follows:

(G1) $T \mid\equiv R \mid\sim r_R$;

(G2) $T \mid\equiv R \mid\sim PID_R$;

(G3) $R \mid\equiv T \mid\sim r_T h$;

(G4) $R \mid\equiv T \mid\sim PID_T$;

(G5) $DB \mid\equiv T \mid\sim PID_T$;

(G6) $T \mid\equiv \# h(PID_R)$;

(G7) $DB \mid\equiv \# h(PID_T)$;

(G8) $T \mid\equiv \# h(k_{ij})$;

(G9) $T \mid\equiv DB \mid\sim k_{ij}$.

The first to the fifth goals show belief requirements. R believes T conveys r_T and PID_T. T believes R conveys r_R and PID_R. DB believes T conveys PID_T. The next two goals show that the messages are not used in the previous sessions and indicate freshness requirements. The eighth goal shows **T** believes that $h(k_{ij})$ is fresh. The last goal indicates T believes DB conveys k_{ij}.

4.4 LOGIC VERIFICATION

Logic verification is based on the assumption, the formalized messages and the related GNY Rules.

From M1, T is informed messages $(PID_R \oplus r_R)$ and $h(PID_R)$. T has not received or sent them in the previous sessions, we have

$$T < *(PID_R \oplus r_R), T < *h(PID_R). \tag{1}$$

Applying the Being-Told Rule T1: $(P \triangleleft (*X))/(P \triangleleft X)$ deduces

$$T < (PID_R \oplus r_R), T < h(PID_R). \tag{2}$$

T can retrieve PID_R from L_R, and applying the Being-Told Rule T2: $(P < X, Y)/(P < X)$ deduces

$$T < PID_R, T < r_R. \tag{3}$$

Thus, T is considered to have been informed r_R and PID_R.

Applying the Possession Rule P1: $(P < X)/(P \ni X)$ deduces

$$T \ni PID_R, T \ni r_R. \tag{4}$$

Applying the Possession Rule P2:

$(P \ni X, P \ni Y)/(P \ni (X, Y), P \ni F(X, Y))$ deduces

$$T \ni (PID_R, r_R). \tag{5}$$

From A4, $T \mid\equiv \# PID_R$, and applying the Freshness Rule F1: $(P \mid\equiv \#(X))/(P \mid\equiv \#(X, Y), P \mid\equiv \# F(X))$ deduces

$$T \mid\equiv \#(PID_R, r_R). \tag{6}$$

From A4, we get

$T \mid\equiv T \xleftarrow{\;PID_R\;} R$.   (7)

Applying the Message Interpretation Rule I3:

$$\frac{P \triangleleft H(X,<S>), P \ni (X,S), P \mid\equiv P \xleftarrow{\;S\;} Q, P \mid\equiv \#(X,S)}{P \mid\equiv Q \mid\sim (X,<S>), P \mid\equiv Q \mid\sim H(X,<S>)}$$

deduces

$T \mid\equiv R \mid\sim (PID_R, r_R)$.   (8)

Finally, from I3 interpretation and applying the Message Interpretation Rule I7:

$(P \mid\equiv Q \mid\sim (X,Y) / P \mid\equiv Q \mid\sim X)$ deduces

$T \mid\equiv R \mid\sim (PID_R)$, $T \mid\equiv R \mid\sim (r_R)$.   (9)

As a result, T believes that R once conveyed PID_R and r_R. Goal G1 and G2 are achieved.

Hereinafter, for simplicity, we directly mark the applied logical rules and Equations behind the Equation.

For Goal 3:

$R < *(r_T \oplus PID_R)$ // by M2   (1)

$R < (r_T \oplus PID_R)$ //by T1   (2)

$R < r_T$ // by T5   (3)

$R \ni r_T$ // by P1   (4)

$R \ni (r_T, PID_R)$ //by P2   (5)

$R \ni H(r_T)$ //by P4   (6)

$R \mid\equiv \varphi(r_T)$ // by R6   (7)

$R \mid\equiv \# PID_R$ //by A6   (8)

$R \mid\equiv \#(r_T, PID_R)$ //by F1   (9)

$R \mid\equiv T \mid\sim r_T$ //by I3   (10)

According to I3, R is entitled to believe that T once conveyed r_T.

For Goal 5:

$DB < *h(PID_T)$, $DB < *PID_R$, $DB < *r_T$ //by M3   (1)

$DB < h(PID_T)$, $DB < PID_R$, $DB < r_T$ //by T1   (2)

$DB \ni PID_R$, $DB \ni h(PID_T)$, $DB \ni r_T$ //by P1   (3)

$DB \ni PID_T$ // by P5   (4)

$DB \ni (r_T, PID_T)$ //by P2   (5)

$DB \mid\equiv \# PID_T$ //by A9   (6)

$DB \mid\equiv \#(r_T, PID_T)$ //by F1   (7)

$DB \mid\equiv T \mid\sim (PID_T)$ // by I3   (8)

According to I3, DB is entitled to believe that T once conveyed PID_T.

For Goal 4:

$R \mid\equiv DB \mid\Rightarrow (DB \mid\equiv *)$ //by A7   (1)

$DB \mid\equiv T \mid\sim (PID_T)$ // by Goal 5   (2)

$R \mid\equiv DB \mid\Rightarrow (T \mid\sim PID_T)$ (3)

$R \mid\equiv DB \mid\equiv (T \mid\sim PID_T)$ //by J3   (4)

$R \mid\equiv (T \mid\sim PID_T)$ //by J1   (5)

As a result, R is entitled to believe that T once conveyed PID_T.

For Goal 6: From message M1 gets

$T < *(PID_R \oplus r_R)$ (1)

$T < (PID_R \oplus r_R)$ //by T1   (2)

$T < PID_R$ //by T2   (3)

$T \ni PID_R$ //by P1   (4)

$T \mid\equiv \# PID_R$ //A4   (5)

$T \mid\equiv \# h(PID_R)$ //by F10   (6)

So T is entitled to believe that $h(PID_R)$ is fresh.

For Goal 7:

$DB \ni PID_T$ //A9   (1)

$DB \mid\equiv \# PID_T$ //A9   (2)

$DB \mid\equiv \# h(PID_T)$ // F10   (3)

So Goal G7 is achieved.

For Goal 8:

$T < *(r_{T_j} \oplus k_{ij}) >> d$ //M5   (1)

$T < (r_{T_j} \oplus k_{ij}) >> d$ // T1   (2)

$T \ni (r_{T_j} \oplus k_{ij}) >> d$ // P1   (3)

$T \ni r_T$ //A1   (4)

So T possesses d, where $d=r_T$. Therefore we get

$T \ni k_{ij}$ (5)

$T \mid\equiv \# k_{ij}$ // A2   (6)

$T \mid\equiv \# h(k_{ij})$ // F10   (7)

As a result, T is entitled to believe that $h(k_{ij})$ is fresh.

For Goal 9: From aforementioned Equation (5) gets

$T \ni k_{ij}$ (1)

Ren Xuping, Zhang Haiping, Lia Yunfa, You Xindong

$$T \ni PID_T , T \models T \xleftarrow{PID_T} DB \quad // \text{ A3} \tag{2}$$

$$T \ni (PID_T , k_{ij}) \quad // \text{ P2} \tag{3}$$

$$T \models \# k_{ij} \quad // \text{ A2} \tag{4}$$

$$T \models \#(k_{ij}, PID_T) \quad // \text{ F1} \tag{5}$$

$$T \models DB |\sim (k_{ij}, PID_T) \quad // \text{ I3} \tag{6}$$

$$T \models DB |\sim (k_{ij}) \quad // \text{ I7} \tag{7}$$

As a consequence, T is entitled to believe that DB once conveyed k_{ij}.

5 Performance analysis

In RFID systems, the performance is another important metric apart from the security issue, so that the optimization and balance between security and performance are necessary for RFID systems [16].

In this section, storage cost, communication load and computation of the proposed protocol are compared with other related protocol. For the sake of simple, KAPP and [11] are chose as the representative of the relevant protocols because of their good performance. The details about performance comparison of [11] can be found in [11]. The details about performance comparison with related protocols of KAPP can be found in [3].

During the entire authentication process of the protocol, each tag and each reader performs one random number generation (RNG) operation and one one-way hash function respectively, while the database performs one one-way hash function like [11]. In KAPP, each tag needs two encryption and two decryption besides generating one random number. Like [3] and [11], access list L_R and L_T are adopted to avoid exhaustive searches in the storage. So computation cost of the protocol is similar to [11] and less than KAPP.

The total authentication progress completed via five phase is acceptable in real sensor system [3]. The communication overhead refers to exchanged messages during each authentication session. In the protocol, there are two exchanged messages between tag and reader during one phase like [11], while KAPP only needs one.

In the protocol, each tag stores access list L_R, the TID ID_T, pseudorandom identifier PID_T and revisable values in the re-writable memory, while other cryptographic algorithms (such as KAAP) also need to store the secret keys. The database stores the role array K_S and the key array K_{MN}, while K_S points out one specific role has own access contents of the tag and K_{MN} indicates the different reader groups own different access authorities. The size of K_{MN} depends on the numbers of the tag groups and the reader groups in the sensor system. A reader group in the proposed protocol can play multiple roles, while a reader group in KAPP can only play one role for one specific tag group. On the whole, the storage requirement of the protocol is less than what it needs in KAPP. Additionally, the memory consumption on one-way hash function is another concern [11]. Standardized cryptographic hash functions such as SHA-1 are too expensive for today's low-cost RFID tags [16,17]. The implementation of the protocol will suitable for low-cost tags.

Above all, the computation cost and the storage requirement of the protocol are less than KAPP, and the communication overhead is similar to [11]. The protocol owns acceptable performance.

6 Conclusions

In the paper, a new distributed protocol is proposed for classified security protection in RFID-based sensor systems. The proposed protocol adopts challenge-response mutual authentication mechanism, random access control mechanism and access lists to strengthen security and privacy protection. As a formal analysis, GNY logic is used to verify the design correctness of the protocol.

Comparison with similar protocol KAPP, the protocol has better scalability because one reader can change its role easily and the tag does not need to store its authentication key. Moreover, the protocol just updates key array on the database side.

According to performance analysis, the protocol has acceptable communication overload, less storage requirements and computation load compared with other related protocol.

Therefore, the proposed protocol is suitable for large-scale RFID application (Such as SCM system). In the future, synchronization problem should be taken into consideration. In addition, RFID authentication protocols with anti-collision mechanism should be paid more attention to.

Acknowledgements

References

[1] Chien H, Chen C 2007 Mutual authentication protocol for RFID conforming to EPC class 1 generation 2 standards *Computer Standards & Interfaces* **29**(2) 254–9
[2] Lee Y, Park Y 2013 A New Privacy-preserving Path Authentication Scheme using RFID for Supply Chain Management *Advances in Electrical and Computer Engineering* **13**(1) 23-6
[3] Ning H, Liu H, Mao J, Zhang Y 2011 Scalable and distributed key array authentication protocol in radio frequency identification-based sensor systems *IET Communications* **5**(12) 1755-68
[4] Sun HM, Ting WC 2009 A *IEEE Trans Mob Comput* **8**(8) 1052-62
[5] Sakai K, Ku W, Zimmermann R, Sun M 2013 *IEEE Trans on computers* **62**(1) 112-23

[6] Wang B, Ma M 2012 *IEEE Trans On Industrial Informatics* **8**(3) 689-96

[7] Piramuthu S 2008 *IEEE Trans Syst Man Cybern C Appl Rev* **38**(3) 360-76

[8] Rahman F, Ahamed S I 2014 Efficient detection of counterfeit products in large-scale RFID systems using batch authentication protocols *Pers Ubiquit Comput* **18** 177-88

[9] Bu K, Liu X, Xiao B 2014 Approaching the time lower bound on cloned-tag identification for large RFID systems *Ad Hoc Networks* **13** 271–81

[10] Morshed Md M, Atkins A, Yu H 2012 Secure ubiquitous authentication protocols for RFID systems *EURASIP Journal on wireless communications and networking* **93**

[11] Ren X, Li Y, Xu X 2013 A one-way Hash Function Based Lightweight Mutual Authentication RFID Protocol *Journal of Computers* **8**(9) 2405-12

[12] Van Deursen T, Mauw S, Radomirovi'c S, Vullers P 2009 Secure ownership and ownership transfer in RFID systems *LNCS* **5789** 637–54

[13] Elkhiyaoui K, Blass E-O, Molva R 2012 ROTIV:RFID Ownership transfer with issuer verification *RFID sec 2011 LNCS* **7055** 163-82

[14] Nan Li, Yi Mu, Willy Susilo, Vijay Varadharajan 2013 Secure RFID Ownership transfer protocols *ISPEC 2013 LNCS* **7863** 189-203

[15] Gong L, Needham R, Yahalom R 1990 Reasoning about belief in cryptographic protocols *IEEE Computer Society Symp Research in Security and Privacy* 234–48

[16] Lehtonen M, Staake T, Michahelles F, Fleisch E 2006 From identification to authentication - a review of RFID product authentication techniques I*n Printed handout of Workshop on RFID Security - RFIDSec 2006*

[17] Weis S 2003 Security and privacy in radio-frequency identification devices *Master's thesis Massachusetts Institute of Technology (MIT) Massachusetts USA*

Authors

Xueping Ren, born in 1978, Zhejiang Province, China

Current position, grades: lecturer at the School of information engineering, Hangzhou Dianzi University.
University studies: B.S. degree from Hanzhou Dianzi University, Hangzhou, China, in 2005.
Scientific interest: RFID.
Publications: 10 papers.

Haiping Zhang, born in 1975, Zhejiang Province, China

Current position, grades: an associate professor at the School of information engineering, Hangzhou Dianzi University.
University studies: B.S. degree from Hanzhou Dianzi University, Hangzhou, China, in 2005.
Scientific interest: on wireless security.
Publications: 10 papers.

Yunfa Li, China

Current position, grades: an associate professor of the School of Computer Science and Technology, Hangzhou Dianzi University.
University studies: PhD degree and the Master degree in Computing Science from Huazhong University of Science and Technology.
Scientific interest: performance modelling, analysis of software, virtual machine, cloud computing, system security, network security.
Publications: over 30 research papers.

Xindong You, China

Current position, grades: a lecturer of School of Computer Science and Technology, Hangzhou Dianzi University, China.
University studies: PhD degree in Northeastern University in 2007.
Scientific interest: distributed computing, cloud storage, energy management.

A new anonymity-based protocol preserving privacy based cloud environment

Jian Wang[1]*, Le Wang[2]

[1]*College of Computer and Information Engineering, Henan University of Economics and Law, China*

[2]*SIAS International University Zhengzhou，China*

Abstract

With the development of cloud computing application, more and more people would like to do business under this environment. But more attention should be paid to the disclosure of privacy during the transaction. Customers will reject to do business on the cloud platform if the cloud environment cannot avoid disclosing their private data. Nowadays, little work has been done about how to prevent sensitive attributes leaking between service providers. Therefore, this paper proposed a new anonymity-based protocol to protect privacy.

Keywords: anonymity, cloud computing, privacy preserving, sensitive attributes, protocol

1 Introduction

Cloud computing raises a range of important privacy issues as acknowledged by a number of recent work [1]. Such issues are due to the fact that, in the cloud, users' data and applications reside on the cloud cluster which is owned and maintained by a third party [2]. Concerns arise since in the cloud it is not clear to individuals why their personal information is requested or how it will be used or passed on to other parties [3, 4].

Despite increased awareness of the privacy issues in the cloud, little work has been done in this space. Recently, Pearson et al. has proposed accountability mechanisms to address privacy concerns of end users [5] and then develop a simple solution, a privacy manager, relying on obfuscation techniques [6]. Their basic idea is that the user's private data is sent to the cloud in an encrypted form, and the processing is done on the encrypted data. The output of the processing is de-obfuscated by the privacy manager to reveal the correct result. However, the privacy manager provides only limited features in that it does not guarantee protection once the data is being disclosed [7-9].

There are many service providers in the cloud, we can call each service as a cloud, each cloud service will exchange data with other cloud, when the data is exchanged between the clouds, and there exists the problem of disclosure of privacy [10]. So my research aims at avoiding the disclosure of the sensitive attributes of users' when user ask for service from the service provider in cloud computing.

2 Related work

There are some privacy problems we want to address in the cloud computing [11]: The first problem is the disclosure of sensitive private information when exchanging data through the cloud service. And the sensitive private information includes: personally identifiable information, usage data, unique device identities and so on [12].

The second problem is that people getting inappropriate or unauthorized access to personal data in the cloud by taking advantage of certain vulnerabilities, such as lack of access control enforcement, security holes and so on [13].

The third problem is that: because the feature of cloud computing is that it is a dynamic environment, in that service interactions can be created in a more dynamic way than traditional e-commerce scenarios [14]. Services can potentially be aggregated and changed dynamically by service providers can change the provisioning of services. In such scenarios, personal sensitive data may move around within an organization or across organizational boundaries, so adequate protection of this information must be maintained despite the changes. So design the method to protect the privacy in cloud computing must meet the dynamical exchange of data.

Nowadays, some researchers focus on cloud data storage security in cloud computing [15]. And other researchers focus on the disclosure of sensitive private information when exchanging data through the cloud computing [16].

Some researchers use identity technology to solve these privacy problems in cloud computing [17]. They propose some requirements for identity services [18]. For

---

*****Corresponding author** e-mail: goodjian121@126.com

example, the identity services should be independent of devices; should permit a single sign-on to thousands of different online services; should allow pseudonyms and multiple discrete identities to protect user privacy; should be interoperable, based on open standards, and should be transparent and auditable.

When create an identity management infrastructure in the cloud computing, we should consider:

1) There are already a number of identity management systems in place on a wide variety of platforms. These need to be supported by the identity management infrastructure. The infrastructure must support cross-system interaction as well as interoperation and delegation between them. This is only possible if the infrastructure and the individual systems are based on open standards, available on all platforms.

2) Identity management systems will support a wide variety of privacy and security properties, ranging from low-security password-based one-factor authentication to high-end, attribute-based systems deploying state-of-the-art privacy-enhancing certificates. While the infrastructure needs to support all of these systems, users should understand the implications of using one system over the other [19].

3 Anonymity-based method

Our paper proposed a new anonymity protocol for the cloud computing services. Before the micro data are published, this anonymity algorithm will process these data. And then send these anonymous data to service providers in the cloud. Then the service provider can integrate the auxiliary information (also called external knowledge, background knowledge, or side information that the service provider can get from other channels such as the web, public records, or domain knowledge) to analyse the anonymous data in order to mine the knowledge they want.

For example, the traffic management board of a region collects records of road accidents for research and analysis. Suppose each record has five attributes, namely occupation, age, vehicle-type, postcode, and faulty. Consider the tuples in Table 1.

The traffic management board anonymised the traffic accident records on attributes age, vehicle-type, and postcode before the data can be released to the service provider in the cloud. Suppose 2-anonymity is required. Table 2 shows a 2-anonymous release of the records with respect to quasi-identifier (age, vehicle-type, postcode).

And then the traffic management board anonymised the traffic accident records on attributes occupation, age, and postcode. Again, suppose 2-anonymity is required. Table 3 shows a 2-anonymous release of the records with respect to quasi-identifier (occupation, age, postcode).

Suppose one service provider in the cloud obtains both releases in Tables 2 and 3. By comparing the two tables, the service provider immediately knows that a family doctor of age 30 driving a white Sedan living in area 31043 was faulty in an accident. The victim may be easily re-

identified by both these anonymous tables.

TABLE 1 A set of traffic accident record

Occupation	Age	Vehicle	Postcode	Faulty
Dentist	30	Red Truck	31043	No
Family doctor	30	White Sedan	31043	Yes
Banker	30	Green Sedan	31043	No
Mortgage broker	30	Black Truck	31043	No

TABLE 2 2-anonymous release of Table 1 with respect to quasi-attributes (age, vehicle-type, postcode)

Occupation	Age	Vehicle	Postcode	Faulty
Dentist	30	Truck	31043	No
Family doctor	30	Sedan	31043	Yes
Banker	30	Sedan	31043	No
Mortgage broker	30	Truck	31043	No

TABLE 3 A 2-anonymous release of Table 1 with respect to quasi-attributes (occupation, age, postcode)

Occupation	Age	Vehicle	Postcode	Faulty
Medical	30	Red Truck	31043	No
Medical	30	White Sedan	31043	Yes
Finance	30	Green Sedan	31043	No
Finance	30	Black Truck	31043	No

Besides, when publish anonymous data, we should consider multiple quasi-identifiers (QI) attributes for different service providers carrying different background knowledge [20]. So the data publishing side should anonymise different quasi-identifiers (QI) attributes of the records for different service providers carrying different background knowledge.

We can take an example to explain. Suppose the traffic management board have records of road accidents for research and analysis. Suppose each record has five attributes, namely occupation, age, vehicle-type, postcode, and faulty. And the traffic management board wants to release these records to different service providers in the cloud.

Suppose there are two service provider named "auto insurance companies" and "human resource department" in the cloud service. Both of them want to use the anonymous records of road accidents.

Importantly, different service providers may carry different background knowledge. For example, the auto insurance company may join the traffic accident records with the vehicle registration records on attributes age, vehicle-type, and postcode to find out whether its customers were at fault in some accidents. Typically, the company does not have the occupation information of its customers, as such information is not required in applying for auto insurance. Therefore, to protect privacy, the traffic management board has to anonymise the traffic accident records on attributes age, vehicle-type, and postcode before the data can be released to the auto insurance company.

Simultaneously, the human resource department may join the traffic accident records with the resident records on attributes occupation, age, and postcode to find out which residents were faulty in some accidents. Therefore, to protect privacy, the traffic management board needs to anonymise the traffic accident records on attributes occupation, age, and postcode. Note that vehicle-type is

not part of the anonymisation, because the human resource department typically does not have information about residents' vehicle types.

Our method is different from the traditional cryptography technology to protect individuals' privacy in the cloud computing services. If we use the cryptography technology. Then we should use cryptography technology to process the data, and send the processed data to service provider in the cloud. Then this service provider can't use these data if it didn't get the key of cryptography. So the service provider in the cloud should use the key to restore the data firstly and then can use these data. However, if we use anonymity technology to process these data and send these anonymous data to service providers in the cloud. Then the service provider can directly use these data without any key and without restoring these data. So this will be more flexible and safe to protect individuals' privacy in the cloud computing services.

4 Private matching protocol

The main purpose for our research is to use private matching technology [21] to intersect user's data and datasets of service provider without accessing each other's data, so this can let the user check whether his anonymous data meet k-anonymity to the datasets of service provider, meanwhile avoiding the disclosure of each other's data.

Our proposed private matching protocol:

1) Both the Client and the SP serialize the attributes and concatenate them to convert their data set into a string of concatenated attribute values, and produce set Vs and set Vc. Which Vc is the set of elements (tuples) in Client, corresponding to the attributes requested by the SP in the current service request, while Vs is the set of elements (tuples) in SP, corresponding to the attributes requested by the SP in the current service request.

Notion 1: Before the user in client adds his tuple to the table of SP, the user should check whether the one new tuple meets k-anonymity to the table of SP. So |Vc|=1. Because before adding one new tuple to the table of SP, the set of Vc only has one tuple to check. But |Vs|>=1. Because the set of Vs may have many tuples before checking.

2) Both Client (named C, for short) and SP (named S, for short) apply hash function h to their sets.

Xc=h (Vc),

Xs=h (Vs).

Each party randomly chooses a secret key. And ec is the key for C, es is the key for S.

3) Both parties encrypt their hashed sets:

Yc=Fec (Xc)=Fec (h (Vc)),

Ys=Fes (Xs)=Fes (h (Vs)).

4) C sends to S its encrypted set

Yc=Fec (h (Vc)).

5a) S ships to C its set Ys=Fes (h (Vs)).

5b) S encrypts each y Yc, with S' key es and sends back to C the pairs

<y, Fes (y)>=<Fec (h (v)), Fes (Fec (h (v)))>.

6) C encrypts each y Ys, with C's key ec, obtaining Zs=Fec (y)=Fec (Fes (h (v))), here the v Vs.

Also, from pairs <Fec (h (v)), Fes (Fec (h (v)))> obtained in Step 5b for the v Vc, It creats pairs <v, Fes (Fec (h (v)))> by replacing Fec (h (v)) with the corresponding v.

7) Because the Vc has only one element v (that is to say Vc has only one tuple). So if this element v Vc meets (Fes (Fec (h (v)))) Zs, then this v Vs. then Vs and Vc has one identical element, else Vs and Vc has no identical element.

Vs and Vc has one identical element is equal to

$$|V_s \cap V_c|=1.$$

Vs and Vc has no identical element is equal to

$$|V_s \cap V_c|=0.$$

5 Case studies

Despite increased awareness of the privacy issues in the cloud, little work has been done in this space. When users order service in the cloud, lots of data including users' attributes need to be transmitted between users and service providers. How to protect users' sensitive attributes and avoid the identification by adversary is still an important problem to solve. Most existing methods are that the user's private data is sent to the cloud in an encrypted form [22], and the processing is done on the encrypted data. However, in this method, the service provider needs to decode these data before they access them. So this will be a waste of time and will be very inconvenient to service providers. Besides, once the data is decrypted the user's privacy may be at risk, since the SP has full control of it. So our proposed approach is different from traditional methods to avoid disclosure of individuals' privacy.

In our proposed approach, the datasets of service provider meet k-anonymity. If a user wants to request a service in the cloud, the user should anonymised his attributes corresponding to the attributes requested by the SP in the current service request, then if this user's anonymised data meet k-anonymity to datasets of this service provider, this user can send his anonymised data to this service provider. Then the service provider will prepare service for this right user. Our proposed approach has two advantages, the first one is the data transmitted in the cloud are anonymised data, so if the adversary get these anonymised data, he cannot identify users. The second advantage is that when service provider gets these anonymised data, service provider can directly use these data without restoring these data. So this will be more flexible and safe to protect individuals' privacy in the cloud.

6 Avoiding privacy indexing

In order to avoid the disclosure of individuals' privacy in cloud, our research address that the user should send the anonymous data to service provider (SP) in the cloud, and the datasets in SP should meet k-anonymity.

The k-anonymity approach publishes a table T' which changes the values on the quasi-identifier attributes so that every tuple in T is published in a group-by of at least k tuples on the quasi-identifier. Publishing T' instead of T can protect privacy effectively against privacy indexing, since the attacker cannot re-identify any individual with a confidence more than 1/k.

In fact the k-anonymity approach meets the second condition of privacy-preserving index (PPI)'s definition in paper [23]. Because the datasets in SP meet k-anonymity, so the datasets in SP also meet the second condition of privacy-preserving index (PPI)'s definition. If an adversary wants to identify who satisfy the query, this adversary executes query on the datasets of SP, then the returned results contain k users (both true positives users and false positives users are in the k users), therefore the probability that the adversary can identify the real user is 1/k. So our based on k-anonymity approach like the method in paper [23], can avoid privacy indexing issue.

Let us take an example.

If an adversary executes a query on the datasets of SP, and this adversary wants to identify user's phone number (sensitive attribute) through this query. Assume table is dataset of one service provider, and table meets 3-anonymity. Then assume the adversary's query is on table.

The query is (select phone number from table where Age=27 & Zip code=555345 & gender=M). So the returned results will be (8142346789, 8123456789, 2346789090, 8456789090). In these results, 8456789090 is true positives and the other three phone numbers are false positives. So the adversary gets four different phone numbers and this adversary can't identify the user whose other attributes satisfy this query. So executing query on these datasets which meet k-anonymity will avoid privacy indexing issue.

Another reason is that the datasets in SP meet 3-anonymity, so this query will return at least 3 different results. All of these results will contain both true positives and false positives. According to the definition of privacy-preserving index (PPI) in paper [23], our proposed approach can avoid privacy indexing issue.

7 Security analysis

Now we use a formal proof method proposed by [24] to prove the security of proposed protocol.

Firstly we give some definitions of the signs used in the proof process.

PD : Private data.

$PDDM$: Privacy-preserving data mining protocol.

PD_{P_i} : Private data of P_i .

EXT_{P_i} : Extra information P_i can obtain through the underlying protocol.

$GAIN_{P_i}$: Advantage of P_i to get access to any other party's private data using a protocol.

$GAIN_{SEC}$: Advantage of P_i to get access to any other party's private data using a protocol by looking at a semantically secure ciphertext which is negligible when using an RSA [25] type of encryption.

$\Pr(PD)$: Probability of disclosing the private data PD without using privacy preserving protocol.

ε : Level of security.

As stated in [24] that if you want to prove whether our proposed protocol named $PRPT$ is privacy preserving, according to any private data, you can find a ε such that:

$$| \Pr(PD \mid PDDM) - \Pr(PD) | \leq \varepsilon .$$

So if we want to prove whether the proposed $PRPT$ protocol is privacy preserving, we only need to find ε such that:

$$| \Pr(PD \mid PRPT) - \Pr(PD) | \leq \varepsilon .$$

Because the advantage of each party P_j by the protocol can access another party's private data can be shown as:

$$GAIN_{P_j} = \Pr(PD_{P_k} \mid EXT_{P_j}, PRPT) - \Pr(PD_{P_k} \mid EXT_{P_j}) ,$$

$(k \neq j) .$

Because each party P_j only runs secure dot product protocol with P_j by using his own randomly generated vector. And party P_j runs secure multi-party addition protocol [26] with other parties by using his private output share. Therefore:

$$GAIN_{P_j} = GAIN_{SEC} , \ (j \neq i) .$$

Because $GAIN_{SEC}$ means that advantage of P_j to get access to any other party's private data, using a protocol by looking at a semantically secure [27] ciphertext which is negligible, when using an RSA type of encryption. So $GAIN_{P_i}$ is also negligible.

Participant P_j by decrypting the message received from other parties and decrypting the signs of their private output, can only know his private value of the final weight vector, which is his own final output. So we can conclude:

$$\varepsilon = \max(GAIN_{P_i}, GAIN_{P_j}) = GAIN_{P_j} .$$

Therefore for each $k, j \in \{1,...,n\}$, $k \neq j$, we can conclude:

$$\Pr(PD_{P_k} \mid EXT_{P_j}, PRPT) - \Pr(PD_{P_k} \mid EXT_{P_j}) \leq GAIN_{P_i} = \varepsilon .$$

So at last we can find $\varepsilon = GAIN_{P_i}$, such that $| \Pr(PD \mid PRRT) - \Pr(PD) | \leq \varepsilon$. Therefore, the proposed $PRPT$ protocol is privacy preserving.

8 Experimental evaluation

In this section we design experiments to show the application and performance of our proposed protocol named *PRPT*. We have used Java language to implement our experiments. The experiments are carried out on Windows XP operating system with 2.13 GHz Intel Core i3 processor and 4 GB of memory. In the following experiments we use IBM Quest Synthetic Data Generator to generate the experimental data.

Figure 1 illustrates that distortion ratio changes depending on the variation of l-value, k-value and i-value. When l-value increases, privacy protection degree increases. The number of different sensitive attributes within the same equivalent group increases, resulting in the generalization level of identifier attributes increased, then leading to substantial information loss from the original data increased, ultimately causing distortion ratio of multiple sensitive attributes set increased. When other conditions are identical, k-value increased, the group size becomes larger. Because the grouped data should meet diverse, attributes in equivalent group increase. Then generalization levels increase and data distortion rate also increases. When the number of data sets and the parameters are equal, the dimension of sensitive attribute i increases, the distortion rate of multiple sensitive attributes is higher, which is caused by sensitive property diversity in each dimension. By contrasting the following figure, we can conclude that *PRPT* protocol by using multiple sensitive attributes generalization can avoid excessive generalization of the identifier attributes, so the total distortion rate is less than the other method. Therefore the total distortion ratio of *PRPT* protocol is smaller than other algorithm. And compared to the l-diversity, k-l-sensitive rules has a relatively low level of data loss. This experiment suggests that *PRPT* protocol has lower information loss than l-diversity algorithm and k-l algorithm. According to the above experimental results analysis, we can conclude that *PRPT* protocol can get

better performance than other algorithm under the same conditions.
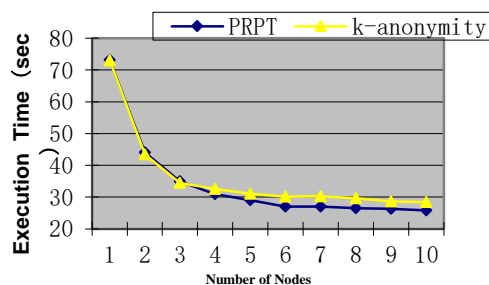
FIGURE 1 Distortion ratio changes depending on l-value changes (k=40, i=3)

9 Conclusions

Cloud computing has been envisioned as the next generation architecture of IT Enterprise. This technology not only gave us more convenience, but also exposed some security problem. When people access cloud service or receive cloud service, they should provide many attributes to ensure this service accomplish successfully. But much private data are included in these attributes. If these sensitive attributes are disclosed, it will bring suffering to people. This paper illustrated the importance of protecting customers' private data in cloud computing. We have argued that it is very important to take privacy into account and we proposed a novel anonymity-based protocol preserving privacy based cloud environment.

Acknowledgements

References

[1] Wang J, , Zhao Y, Shuo J, Le J 2009 Providing Privacy Preserving in Cloud Computing *in ICTM Proceedings IEEE* 213-6

[2] Almorsy M 2011 Collaboration-Based Cloud Computing Security Management Framework, *in CLOUD Proceedings IEEE* 364-71

[3] Kretzschmar M, Golling M 2011 Security management spectrum in future multi-provider Inter-Cloud environments-Method to highlight necessary further development *in SVM Proceedings IEEE* 1-8

[4] Bouayad A, Blilat A, El Houda Mejhed N, El Ghazi M 2012 Cloud computing: Security challenges *in CIST Proceedings IEEE* 26-31

[5] Khalil I, Khreishah A, Bouktif S, Ahmad A 2013 Security Concerns in Cloud Computing *in ITNG Proceedings IEEE* 411-6

[6] Kandukuri B, Paturi V, Rakshit A 2009 Cloud Security Issues *in SCC Proceedings IEEE* 517-20

[7] Hamdi M 2012 Security of cloud computing, storage, and networking *in CTS Proceedings IEEE* 1-5

[8] Behl A, Behl K 2012 Security Paradigms for Cloud Computing *in CICSYN Proceedings IEEE* 200-5

[9] Al Awadhi E, Salah K, Martin T 2013 Assessing the security of the cloud environment *in GCC Proceedings IEEE* 251-6

[10] Wang J, Le J A New Privacy Preserving Approach Used in Cloud Computing *Journal of Key Engineering Materials* **439-440**(1) 1318-23 *(in Chinese)*

[11] Francis T, Vadivel S 2012 Cloud computing security: Concerns, strategies and best practices *in ICCCTAM Proceedings IEEE* 205-7

[12] Wang J, Novel 2012 A Anonymity Algorithm for Privacy Preserving in Publishing Multiple Sensitive Attributes *Research Journal of Applied Sciences, Engineering and Technology* **4**(22) 4923-7

[13] Mell P 2012 What's Special about Cloud Security *Journal of IT Professional* **14**(4) 6-8

[14] Viega J 2012 Cloud Security: Not a Problem *Journal of Security & Privacy* **10**(4) 3

[15] Goth G 2011 Public Sector Clouds Beginning to Blossom: Efficiency, New Culture Trumping Security Fears *Journal of Internet Computing* **15**(6) 7-9

[16] Kaufman L M 2009 Data security in the World of Cloud Computing *in Security and Privacy Proceedings IEEE* 61-4

[17] Behl A, Behl K 2012 An analysis of cloud computing security issues *in WICT Proceedings IEEE* 109-14

Computer and Information Technologies

[18] Wang J, Le J 2010 Novel Privacy Preserving Classification Mining Approach Based Cloud Environment *in BTMC Proceedings IEEE* 236-9

[19] Wang J, Luo Y, Zhao Y, Le J 2009 A Survey on Privacy Preserving Data Mining *in DBTA Proceedings IEEE* 111-4

[20] Zhang Q, Koudas N, Srivastava D, Yu T 2007 Aggregate query answering on anonymized tables *in ICDE Proceedings IEEE* 116-25

[21] Li Y, Tygar JD, Hellerstein J M 2005 Private Matching, in Computer Security *Proceedings Springer* 25-50

[22] Siani P, Yun S, Miranda M 2009 A privacy manager for cloud computing *In CloudCom Proceedings IEEE* 90-106

[23] Mayank B, Rakesh A, Jaideep V 2009 Privacy-preserving indexing of documents on the network *The VLDB Journal* 837-56 *(in Chinese)*

[24] Zhan J, Matwin S 2006 A Crypto-Based Approach to Privacy-Preserving Collaborative Data Mining *in Data Mining Proceedings IEEE* 546-50

[25] Wang S, Liu G 2011 File Encryption and Decryption System Based on RSA Algorithm *in Computational and Information Sciences Proceedings IEEE* 797-800

[26] Liu W, Wang Y, Cao Y 2011 Research on Secure Multi-Party Closest String Problem *in Wireless Communications, Networking and Mobile Computing Proceedings IEEE* 1-4

[27] Lee J, Chang J 2007 Semantically Secure Authenticated Encryption Scheme and the Same Scheme for Ad-hoc Group Called a Ring *in Information Technology Proceedings IEEE* 825-30

Authors

Jian Wang, born in 1981, November 26th, China

Current position: an instructor in Henan University of Economics and Law.
University studies: Ph.D. degree from Donghua University, Shanghai, China in June 2011.
Scientific interest: privacy preserving in cloud computing, information security.
Publications: 10 papers.

Le Wang, born in 1983, October 29th, China

Current position: an instructor in SIAS International University.
University studies: Bachelor degree from Zhengzhou University, Zhengzhou, China in June 2005.
Scientific interest: computer science, database application.

The research of data conflict in digital camp management and control system based on IOT

Xingchen Li[1*], Shenglin Li[2], Heng Zhang[2], Hui Cai[1]

[1]*Graduate school, Logistical Engineering University of P.L.A, Chongqing, China*

[2]*Department of Information Engineering, Logistical Engineering University of P.L.A, Chongqing, China*

Abstract

A core technology in IOT (Internet of Thing) is data processing and there may be a mismatch between data collection and data interface. Aiming at data conflicts in the DCMCS (Digital Camp Management Control System), this paper defines data conflicts with formal description and divides the data conflicts of the DCMCS into six types. For each type of the data conflicts, a resolution strategy is designed to solve it. Then the paper designs the DCRS (Data Conflict Resolution System) to detect and resolve the data conflicts automatically and quickly. The solutions provide technical support for data interaction, which have been successfully verified in the DCMCS.

Keywords: formal description, data conflict, resolution strategy, DCRS

1 Introduction

As an important goal of modern barracks to transform information, the DCMCS (Digital Camp Management Control System) refers to utilize sensing technology, network communications and automatic control technology to build IOT with barracks facilities and equipment, getting the operational status of devices in real time and controlling them accurately. In the process of collecting data, there may be a mismatch between data acquisition and the upper data interface due to sensor noise, external interference, device aging or deviation of application itself and resultant data conflicts. Therefore, it is important to research data conflicts in the DCMCS and resolution method.

2 Backgrounds

A core technology in IOT is data processing. Data processing in IOT can be divided into data recognition and data interaction [1]. DCMCS data conflict is the data exchange errors arising from mismatches between data interface and device identification, data types, data structures, data accuracy and acquisition space. That is, when output data of sensing devices cannot meet the restrictions of the data input interface, it will produce data conflict. For example, sensor D1 outputs float-type data and integer-type is required by the interface D2, so it will generate data type conflict.

Current research of data Processing in IOT is about heterogeneous data storage and abnormal data detection. In terms of storage and query processing of sensor sampling data, distributed storage method [2-4] means that sample data are stored directly in each sensor node or in

the sink node, and are obtained through remote access query processing. If query processing involves different sensor networks, it needs to be converted by the middleware to achieve interoperability operation between heterogeneous sensor networks [5, 6]. With respect to centralized processing of sensor data, the most direct way is to use cloud data management and related technologies. Currently a majority of cloud data management systems are the "key-value" database such as Bigtable [7], Dynamo [8], Hbase [9] and so on. "Key-value" database can efficiently handle queries based on the primary key, but fails to support networked data storage, spatial and temporal logic query and attribute constraints inquiries. Parallel database technology [10, 11] supports the processing of structured data by organizing multiple relational databases into a massive database cluster, but the method cannot retrieve quickly the required data based on the identity of the sensor. Common methods of anomaly data detection are statistics, feature selection, neural networks, data mining and wavelet singularity detection. The main idea based on statistics [12] is to assume that data sets obey certain distribution or probability model and that serious deviation from the distribution curve object are recorded as outliers by calculating the probability of the object in the distribution model. Outlier based on distance was first proposed by the Knor and Ng [13], who viewed the records as points in data space. Outlier is defined as the distance between the point and most points of data sets, which are larger than the value of a certain point.

Literature reveals that credibility or a statistical model would normally be added in the uncertain data of sensors to detect abnormal data. Data conflicts due to a mismatch between sensor data acquisition and the interface need to be researched further. In view of this, the paper proposes a

*Corresponding author e-mail: since0701@163.com

data conflict detection method based on formal description, and then puts forward a conflict resolution program to quickly determine and digest data conflict in the environment of mass data.

3 Formal description of data conflict

Formal description is the method of describing the system properties with the mathematical basis formal symbol for the specification, design and validation of computer systems. It can accurately and succinctly describe phenomena, objects or action results and is ideal for modelling tools. Samples, data conflicts, data interfaces and interface matching principles are described and defined in this section. The accurately formal description of data conflicts can provide good support for the digestion of data conflict and provide a basis for the development of the DCRS.

The DCMCS needs to read the device identification and sampling data, including sampling value, and spatial and temporal information, such as sampling time and sampling location, which reflect the characteristics of the target object. Sample sequence can be composed of a number of samples of the device. Due to the heterogeneity of collection data, a uniform representation is needed for samples which sensing devices collected with unique identity, time, location, sample values and units of measurement.

Definition 1 (The Definition of Sample-Value): In the DCMCS, Sample-Value is expressed as $S = <ID, T, L, V, U>$, where ID is the unique identification of sampling device; T is the acquisition time; L represents the sampling location. For a stationary device, sampling location is the place that the sensor is installed and for mobile devices, it means the places where collect action takes place; V is the value of sensor data collection; U is the measure unit of data value. S_i shows the sample data of ith device.

After data collection, the sensor needs to exchange the data with the data management system through data interface to process data collected from the sensors.

Definition 2: (The Definition of Data Interface) In the DCMCS, the data interface of the data management system is expressed in a ternary suit $A = <I, G, D>$, $I = \{I_1, I_2 \cdots I_n\}, (n \in N)$ indicate the function set of the data interface. $I_i (i \in N, 1 \leq i \leq n)$ show the ith interface of the data interface; $G = \{G_1, G_2 \cdots G_n\}, (n \in N)$ indicate the functional function set of the data interface, G_i show the set of functional function corresponding with interface I_i ; $D = \{D_1, D_2 \cdots D_n\}, (n \in N)$ indicate the set of input and output data of the data interface, and D_i show the set of input and output data corresponding with interface I_i .

Definition 3: (The Definition of Input/Output of Data Interface) In the data interaction, the data interface $I = <I^{in}, I^{out}>$, $I^{in} = \{P_1, P_2 \cdots P_m\}, (m \in N)$ indicate the

input of data interface. $I^{out} = \{P_1, P_2 \cdots P_n\}, (n \in N)$ indicate the output of data interface, and $P = <M, F, J, R>$ show the parameters of input/output and restriction conditions of those parameters, where M indicate the types of input/output parameters, F indicate the structure of input/output parameters, J indicate the precision of input/output data, R indicate the value range of input/output data. $P_i (i \in n)$ indicates the ith parameter and its restriction conditions of input/output interface I.

Definition 4: (The Definition of Data Interface Matching) Interface $A_i = <I_i, G_i, D_i>$, where $I_i = <I_i^{in}, I_i^{out}>$, the input of interface $I_i^{in} = P_i^{in} = <M_i^{in}, F_i^{in}, J_i^{in}, R_i^{in}>$, the output of sensor $I_i^{out} = P_i^{out} = <M_i^{out}, F_i^{out}, J_i^{out}, R_i^{out}>$, if $\exists (P_i^{out} = P_i^{in})(i \in n)$, output data I^{out} matches input data I^{in} , denoted as $I^{out} \Rightarrow I^{in}$.

Definition 5: (The Definition of Data Conflict) Define

$$H_{A_i} = \begin{cases} 1, A_i \ Normal \ \text{int} eraction \\ 0, A_i \ Abnormal \ \text{int} eraction \end{cases}, (i \in N) \quad . \quad \text{In the}$$

DCMCS $\exists A_i$, if $P_i^{out} \neq P_i^{in}$, $I^{out} \Rightarrow I^{in}$ is not established, we call it 'data conflict' and $H_{A_i} = 0 (i \in N)$ at this time.

4 Types of data conflict

Data conflicts of the DCMCS based on IOT are mainly caused by the identity of the sensor, the fact that the output data does not match the constraints of the interface and so forth. Data conflicts can be summarized as the following six types: ID conflict, data type conflict, data structure conflict, data precision conflict, data overflow conflict and acquisition space conflict, as shown in Figure 1.

Type of Data Conflict	
ID conflict	data accuracy conflict
data type conflict	data overflow conflict
data structure conflict	acquisition space conflict
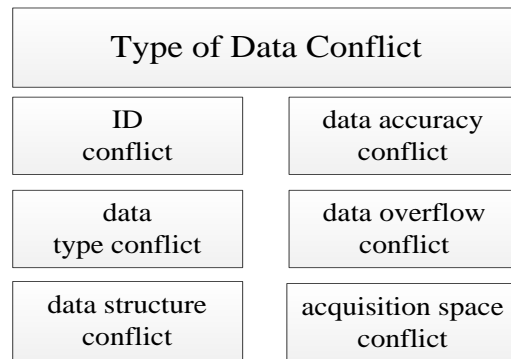
FIGURE 1 The type of data conflict

ID conflict: The IOT of digital camps has a lot of sensor nodes and it needs to encode these sensing devices and give them unique ID for identification. During the dynamic process of adding or deleting devices, it may encode different sensors with the same ID. The system cannot locate the unique sensing device and then generates data conflict. In a single round of data collection process, $\exists A_i = <I_i, G_i, D_i>$, $\exists A_j = <I_j, G_j, D_j>$, where $D_i = <ID_i, T_i, L_i, V_i, U_i>$, $D_j = <ID_j, T_j, L_j, V_j, U_j>$. If $ID_i =$

ID_j, we call it *ID* conflict and denote as $Conf_{ID}(IDi, IDj)$. *ID* conflict occurs when samples that have the same *ID* try to join sample sequences, causing the collect data not be received normally by the data interface.

Data type conflict: As it cannot transfer impliedly in data interaction, data type conflict will occur if the data types of input and output are inconsistent and lead to errors in data exchange. $\exists A_i =< I_i, G_i, D_i >$, where $I_i =< I_i^{in}, I_i^{out} >$, $I_i^{in} = P_i^{in} =< M_i^{in}, F_i^{in}, J_i^{in}, C_i^{in} >$, $I_i^{out} = P_i^{out} =< M_i^{out}, F_i^{out}, J_i^{out}, C_i^{out} >$, if $M_i^{in} \neq M_i^{out}$, then $P_i^{in} \neq P_i^{out}$, that means $I^{out} \Rightarrow I^{in}$ is not established and makes $H_{A_i} = 0$. We call the data conflict as data type conflict and denoted as $Conf_{type}(DType1, DType2)$.

Data structure conflict: Data structure is a kind of type set and it is a plurality of attributes including the number of parameters and data types. *DStruct1* and *DStruct2* could be the structure types defined by system or itself, which represent the output data structure and the input data structure. If there is an inconsistent type in *DStruct1* and *DStruct2*, data structure conflict appears. The interface cannot recognize or accept the data collected. $\exists A_i =< I_i, G_i, D_i >$, where $I_i =< I_i^{in}, I_i^{out} >$, $I_i^{in} = P_i^{in} =< M_i^{in}, F_i^{in}, J_i^{in}, C_i^{in} >$, $I_i^{out} = P_i^{out} =< M_i^{out}, F_i^{out}, J_i^{out}, C_i^{out} >$, if $F_i^{in} \neq F_i^{out}$, then $P_i^{in} \neq P_i^{out}$, that means $I^{out} \Rightarrow I^{in}$ is not established and makes $H_{A_i} = 0$. We call it data structure conflict, denoted as $Conf_{struct}(DStruct1, DStruct2)$.

Data accuracy conflict: Data accuracy will directly affect the result of data calculation. If the precision of input data and output data is inconsistent, for example the output data of sensor is float while the interface needs integer data, a data conflict generates and the interface cannot get correct data. $\exists A_i =< I_i, G_i, D_i >$, where $I_i =< I_i^{in}, I_i^{out} >$, $I_i^{in} = P_i^{in} =< M_i^{in}, F_i^{in}, J_i^{in}, C_i^{in} >$, $I_i^{out} = P_i^{out} =< M_i^{out}, F_i^{out}, J_i^{out}, C_i^{out} >$, if $J_i^{in} \neq J_i^{out}$, then $P_i^{in} \neq P_i^{out}$, that means $I^{out} \Rightarrow I^{in}$ is not established and makes $H_{A_i} = 0$. We call the data conflict as accuracy conflict, denoted as $Conf_{Accur}(DAccur1, DAccur2)$.

Data overflow conflict: In the data exchange process, sensor collecting data exceeding the defined range of data interface and generating value range conflict is a kind of normal phenomenon. For example, the original collecting data of pressure sensor is 42560 and the data interface is defined as 0-42559 corresponding to a pressure of 0-1Mpa, then data conflict occur. $\exists A_i =< I_i, G_i, D_i >$, where $I_i =< I_i^{in}, I_i^{out} >$, $I_i^{in} = P_i^{in} =< M_i^{in}, F_i^{in}, J_i^{in}, C_i^{in} >$, $I_i^{out} = P_i^{out} =< M_i^{out}, F_i^{out}, J_i^{out}, C_i^{out} >$, $D_i =< ID_i, T_i, L_i, V_i, U_i >$, if $V_i \notin R_i$, then $P_i^{in} \neq P_i^{out}$, that means $I^{out} \Rightarrow I^{in}$ is not established and makes $H_{A_i} = 0$.

We call the data conflict as data overflow conflict, denoted as $Conf_{range}(DRange1, DRange2)$.

Acquisition space conflict: Due to the spatial and temporal characteristics of the data collected, sensors can only collect data in one place at the same time. In a single round of the data collection process, $\exists A_i =< I_i, G_i, D_i >$, $\exists A_j =< I_j, G_j, D_j >$, where $D_i =< ID_i, T_i, L_i, V_i, U_i >$, $D_j =< ID_j, T_j, L_j, V_j, U_j >$. If $L_i = L_j$, we call the data conflict as acquisition space conflict, denoted as $Conf_{Loc}(Loc1, Loc2)$.

5 Strategies to resolve data conflict

On basis of formal description, we divide data conflicts into six categories: *ID* conflict, data type conflicts, data structure conflict, data accuracy conflict, data overflow conflict and acquisition space conflict. Hence, there are also six strategies to resolve data conflicts, as shown in this section.

Digestion strategies for *ID* conflict: When an *ID* conflict occurs, rename the sensor according to the naming rule that the acquisition time goes behind the other sensor, in order to make the identity different. The implementation procedures are, first, to determine the identity of the two sensors. If they are inconsistent, data can be directly released; otherwise compare the acquisition time of samples T_1 and T_2. If $T_1 > T_2$, then rename ID_1 according to the naming rule; if $T_1 < T_2$, then rename ID_2. The realization function is: private SensorData ReID(SensorData data1) Figure 2:

```
1  public class SensorData
2      {
3      private string _id;
4      public string Id
5      {
6      get { return _id; }
7      set { _id = value; }
8      }
9      private DateTime _measuringTime
10     public DateTime MeasuringTime
11     {
12     get { return _measuringTime; }
13     set { _measuringTime = value; }
14     }
15     }
16     private SensorData ReID(SensorData data1)
17     {
18     int count = 0;
19     SensorData ReSensorData = new SensorData();
20     ReSensorData = data1;
21     foreach (SensorData data in sensorDatas)
22     {
23     if (data.Id.Equals(data1.Id))
24     count++;
25     if (count > 1 && (data.MeasuringTime >
       data1.MeasuringTime))
26     ReSensorData = data;
27     }
28     return ReSensorData;
29  }
```

FIGURE 2 The source code of digestion strategies for *ID* conflict

Digestion Strategies for data type conflict: Converting the data type of the sensor output to make it convert with

the input of interface. The implementation procedures are, first, to judge the consistency of output data type $DType_1$ and interface data type $DType_2$. If they are consistent, data can be directly released; otherwise, it needs to transfer the type of sensor output data $Dtype_1$ according to the type of interface data. So that it can meet the constraints of the interface, like Integer transfer to String or String transfer to Integer, since the data interfaces are public and can be realized with middleware, database management systems and other forms. The realization function is: private SensorData ReDataType(SensorData data1) Figure 3:

```
1 public class SensorData
2 {
3   private string _dataType;
4   public string DataType
5   {
6   get { return _dataType; }
7   set { _dataType = value; }
8   }
9   private string _dataValue;
10  public string DataValue
11  {
12  get { return _dataValue; }
13  set { _dataValue = value; }
15  }
16  }
17  private SensorData ReDataType
          (SensorData data1)
18  {
19  SensorData ReSensorData = new SensorData();
20  Boolean success = true;
21  try
22  {
23   switch (data1.DataType)
24  {
25  case "int":
26   Convert.ToInt32(data1.DataValue); break;
27   case "float":
28  double.Parse(data1.DataValue); break;
29   case "date":
30  Convert.ToDateTime(data1.DataValue); break;
31   default: break;
32  }
33  }
34  catch
35  {
36  success = false;
37  }
38  if (success)
39  return ReSensorData;
40  else
41  {
42  return chooseOtherDataValue(data1.Id);
43  }
44  }
```

FIGURE 3 The source code of digestion strategies for data type conflict

Digestion Strategies for data structure conflict: Converting structure of the sensor output data in order to make it convert with the structure of interface input data. The implementation procedures are, first, to compare the structure of the sensor output data $Dstructure_1$ with the structure of interface input data $Dstructure_2$; analyse the consistency of the structures of two samples, including the number of acquisition parameters, data type and other attributes. Then transfer $Dstructure_1$ consistent with $Dstructure_2$ which is not mismatched to make it meet the constraints of the interface. The realization function is:

private SensorData ReDataStruct(SensorData data1) Figure 4:

```
1 public class SensorData
2 {
3   private string _dataStruct;
4   public string DataStruct
5   {
6   get { return _dataStruct; }
7   set { _dataStruct = value; }
8   }
9   private string _address;
10  public string Address
11  {
12  get { return _address; }
13  set { _address = value; }
14  }
15  }
16  private SnsorData ReDataStruct(SensorData data1)
17  {
18  if (!data1.DataStruct.Split('-')[1].Equals("1"))
19  {
20  IList<int> count=new  List<int>();
21  foreach (SensorData data in sensorDatas)
22  {
23  string[] numbers = data1.DataStruct.Split('-');
24  if (data.Address.Equals(data1.Address))
25  {
26  count.Add(int.Parse(numbers[0]));
27  }
28  }
29  count.Add(int.Parse(data1.DataStruct.
    Split('-')[0]));
30  bool repeat=true;
31  if (count.Count == int.Parse(data1.DataStruct.Split('-')[1]))
32  {
33  for (int i = 0; i < count.Count - 1; i++)
34  {
35  for (int j = 0; j < count.Count - 1; j++)
36  {
37  if (count[i] == count[j] && i != j)
38   repeat = true;
39  }
40  }
41  if(repeat)
42  {
43  return chooseOtherDataValue(data1.Id);
44  }
45   else return data1;
46  }
47   else
48  {
49  return chooseOtherDataValue(data1.Id);
50  }
51  }
52  else
53  return data1;
54  }
```

FIGURE 4 The source code of digestion strategies for data structure conflict

Digestion Strategies for data accuracy conflict: The digestion strategies for data accuracy conflict are to unite the accuracy of sensor output data and the accuracy of interface input data with the same constraints. The implementation procedures are to calibrate $DAccuracy_1$ according to the accuracy of interface $DAccuracy_2$. For values larger than accuracy, use the rounded method and for those smaller than accuracy, use the padding method at the end of zero. The realization function is: private SensorData ReDataAccuracy(SensorData data1) Figure 5:

```
1 public class SensorData
2        {
3            private string _dataAccuracy;
4            public string DataAccuracy
5            {
6            get { return _dataAccuracy; }
7            set { _dataAccuracy = value; }
8            }
9            private string _dataValue;
10           public string DataValue
11           {
12           get { return _dataValue; }
13           set { _dataValue = value; }
14           }
15           }
16       private SensorData ReDataAccuracy(SensorData data1)
17           {
18           if (data1.DataAccuracy.Equals(""))
19           return data1;
20           else
21            {
22           if (data1.DataAccuracy.Contains('.'))
23           {
24           int accuracy =
25           Convert.ToInt32(data1.
           DataAccuracy.Split('.')[0]);
26           if (data1.DataValue.Length <= accuracy)
27           {
28           int accuracy1 = Convert.ToInt32(data1.
           DataAccuracy.Split('.')[1]);
29           if (data1.DataValue.Length <= accuracy1)
30           {
31           return data1;
32           }
33           else
34           {
35           return chooseOtherDataValue(data1.Id);
36           }
37           }
38           else
39           {
40           return chooseOtherDataValue(data1.Id);
41           }
42           }
43           else
44           {
45           int accuracy = 31Convert.ToInt32(data1.
           DataAccuracy);
46           if (data1.DataValue.Length <= accuracy)
47           {
48           return data1;
49           }
50           else
51           {
52            return chooseOtherDataValue(data1.Id);
53           }
54           }
55           }
56 }
```

FIGURE 5 The source code of digestion strategies for data accuracy conflict

Digestion Strategies for data overflow conflict: The digestion strategies for data overflow conflict are to limit the value range of input/output $DRange_1$ and $DRange_2$ to exclude the overflow data. The implementation procedures are that the value range of input/output is $DRange1 \cap DRange2$. If $DRange1 \cap DRange2 = \emptyset$, the sample can't interact with the interface. System alarms to remind invalid data and the sensor output value range needs to be defined so that it does not exceed the value

range. The realization function is: private SensorData ReDataRange(SensorData data1) Figure 6:

```
1 public class SensorData
2        {
3            private string _dataValue;
4            public string DataValue
5            {
6            get { return _dataValue; }
7            set { _dataValue = value; }
8            }
9            private string _minData;
10           public string MinData
11           {
12           get { return _minData; }
13           set { _minData = value; }
14           }
15           private string _maxData;
16           public string MaxData
17           {
18           get { return _maxData; }
19           set { _maxData = value; }
20           }
21           }
22       private SensorData ReDataRange(SensorData data1)
23           {
24       if (data1.DataType.Equals("int") ||
         data1.DataType.Equals("float"))
25           {
26       if (double.Parse(data1.DataValue) >=
         double.Parse(data1.MinData) &&
         double.Parse(data1.DataValue) <=
         double.Parse(data1.MaxData))
27           {
28       return data1;
29           }
30       else
31           {
32       alert();
33       return chooseOtherDataValue(data1.Id);
34           }
35           }
36       else return data1;
37           }
```

FIGURE 6 The source code of digestion strategies for data overflow conflict

Digestion Strategies for acquisition space conflict: The digestion strategies for acquisition space conflict are to correct the data acquisition location. The implementation procedures are to determine the type of the sensor when an acquisition space conflict occurs. If it is a mobile device, then release the data; For a fixed device, modify the error collection address according to the device installation table that records the installation site. Its function is: private SensorDataReDataLocation (SensorData data1) and private SensorData chooseOtherDataValue(string id) Figure 7:

```
1 public class SensorData
2        {
3            private int _deviceType;
4            public int DeviceType
5            {
6            get { return _deviceType; }
7            set { _deviceType = value; }
8            }
9            private string _location;
10           public string Location
11           {
```

```
12          get { return _location; }
13          set { _location = value; }
14          }
15          }
16          private SensorData ReDataLocation
            (SensorData data1)
17          {
18          if (data1.DeviceType.Equals(0))
19          {
20          if(data1.Location.Equal
            (getLocation(data1.Id)))
21          {
22          return data1;
23          }
24          else
25          {
26          return chooseOtherDataValue(data1.Id);
27          }
28          }
29          else return data1;
30          }
31          private SensorData choose
            OtherDataValue(string id)
32          {
33          return new SensorData();
34          }
```

FIGURE 7 The source code of digestion strategies for acquisition space conflict

6 Realization of data conflict resolution system

On the basis of the formal description of data conflict and strategies to solve data conflict, the paper designs the Data Conflict Resolution System in this section to detect and solve the data conflict automatically and quickly. DCRS aims at adopting relevant strategies for each sort of data conflict to realize the goal of resolving conflicts while at the same time increasing system stability and reliability. If the input and output of data interface does not match, the DCRS calls the conflict resolution to deal with invalid data. Its framework flow is showed in Figure 8.
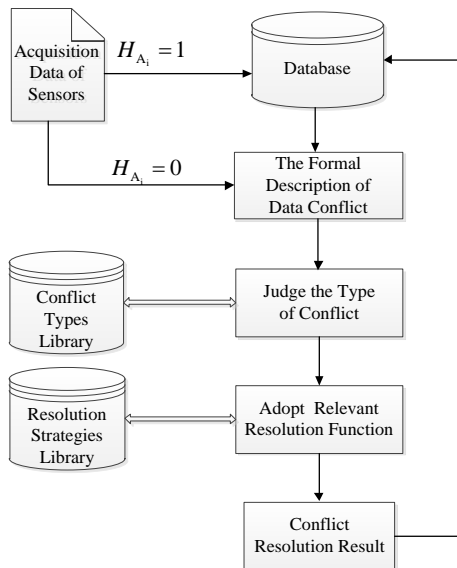
FIGURE 8 The framework flow of DCRS

In the framework flow of the DCRS, conflict type library has six data conflicts and resolution strategies library has resolution function corresponding to the

conflicts. Firstly, the DCRS detects data conflicts in the input and output of the interface. If the samples match the constraints of the interface, the DCRS releases data into database; otherwise，the DCRS identifies the type of the data conflict according to the formal description of the conflict type library, and then calls the appropriate function to resolve conflicts. After that, the DCRS releases the data into database.

The DCRS adopts Microsoft. NET platform for developing and data conflict digestion algorithm uses the technology of C# language. Visual Studio. NET is Microsoft's second-generation development tool for building and deploying powerful and secure software. The database system of the DCRS uses Oracle 10g, the relational database management system with features of large data capacity, persistent storage time, convenient data sharing, and high reliability. The database system is used to store the conflict type and dynamic data generated during the system operation. The main interface of the DCRS is shown in Figure 9.

FIGURE 9 The main interface of DCRS

There are a variety of criteria to evaluate algorithm performance, and what kind of standard is adopted is mainly depended on the system requirements. In the DCMS real-time is the most important factor, and therefore the processing time overhead and memory space overhead can be used to evaluate the performance of the data conflict resolution algorithm. Processing time overhead is the time that digestion algorithm takes to process the data conflict. Transaction will be worthless to the system if it is completed after the zero-value point and is a waste of system resources. Memory space overhead is the memory occupied by the algorithm when it digests the data conflicts and is used to store the information generated in the process. Algorithm that is rational designed should effectively utilize system resources to process the data conflict without taking up too many resources. The PC that the DCRS runs on with CPU clock at 2.5GHz and 4GB memory is used to test the algorithm performance. The performance test uses one hundred thousand data collected by the water meters and electric meters in the DCMCS. There are 181 abnormal data when

Li Xingchen, Li Shenglin, Zhang Heng, Cai Hui

the test finishes and it costs 2505ms to resolute data conflicts. Among the data conflicts, there are 7 ID conflicts that need 31.5ms to process, 4.5ms for each conflict. There are 35 data type conflicts that need 318.5ms to process, 9.1ms each. There are 28 data structure conflicts that need 336ms to process, 12ms each. There are 53 data accuracy conflicts that need 726.1ms to process, 13.7ms each. There are 46 data overflow conflicts that need 230ms to process, 5ms each. There are 12 acquisition space conflict that need 75.6ms to process, 5.3ms each. The processing time overhead for the six types of conflict digestion algorithm is showed in Figure 10.
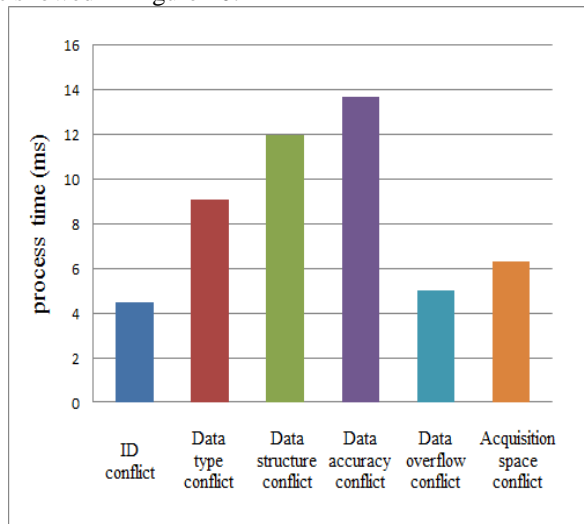
FIGURE 10 The processing time overhead for the six types of conflict digestion algorithm

The memory space overhead for the six types of conflict digestion algorithm is shown in Figure 11. It can be seen from the test that the digestion algorithm can meet the real-time data processing system requirements.
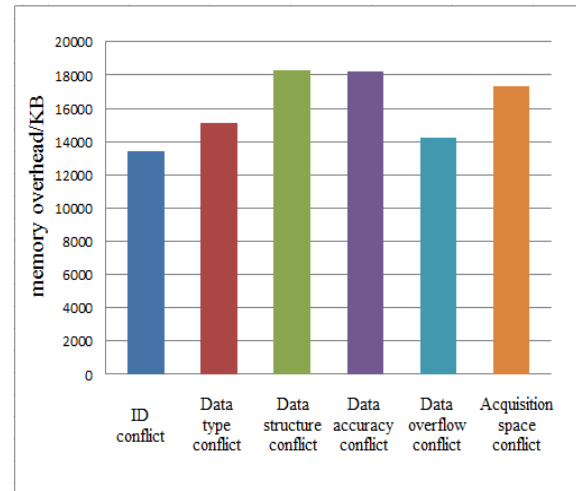
FIGURE 11 The memory space overhead for the six types of conflict digestion algorithm

7 Conclusions

In the DCMCS, there are data conflicts caused by a lack of uniform standards for interaction between the device of IOT and the upper data interface. To address this problem, the paper analyses formal description of data conflict, proposes strategies to resolve conflict, designs resolution flow and exploits the DCRS. The solutions have been successfully verified in the DCMCS, realizing intellectualized resolution of data conflict and providing technical support and insightful guarantee for data interaction.

Acknowledgment

References

[1] Hu Y L, Sun Y F, Yin B C 2012 Information sensing and interaction technology in Internet of Thing *China Journal of Computers* **35**(6) 1147-63 *(in Chinese)*

[2] Tsiftes N, Dunkels A 2011 A database in every sensor *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems Seattle USA* 316-29

[3] Chen X, He X, Guo H, Wang Y 2011 Design and Evaluation of an Online Anomaly Detector for Distributed Storage Systems *Journal of Software* **6**(12) 2379-90 *(in Chinese)*

[4] Madden S, Franklin M J, Hellerstein J M, Hong W, Tiny 2005 DB: An acquisitional query processing system for sensor networks *ACM Transactions on Database Systems* **30**(1) 122-73

[5] Gurgen L, Roncancio C, Labbé C 2008 SStreaMWare: A service oriented middleware for heterogeneous sensor data management *Proceedings of the 5th International Conference on Pervasive Services (ICPS'08) Sorrento Italy* 121-30

[6] Kim M, Lee J W, Ryou J C 2008 COSMOS: A middleware for integrated data processing over heterogeneous sensor networks. *ETRI Journal*, **30**(5) 696-706

[7] Chang F, Dean J, Ghemawat S 2006 Bigtable: A distributed storage system for structured data *Proceedings of the 7th Symposium on Operating Systems Design and Implementation (OSDI'06) Seattle USA* 205-8

[8] DeCandia G, Hastorun D, Jampani M 2007 Dynamo: Amazon's highly available key-value store *Proceedings of the 21st ACM Symposium on Operating Systems Principles (SOSP'2007) Washington USA* 205-20

[9] Carstoiu D, Lepadatu E, Gaspar M 2010 Hbase: Non SQL database, performances evaluation *International Journal of Advancements in Computing Technology* **12**(5) 42-52

[10] Zhou Y, Zhu Q, Zhang Y 2011 Spatial Data Dynamic Balancing Distribution Method Based on the Minimum Spatial Proximity for Parallel Spatial Database *Journal of Software* **6**(7) 1337-44

[11] Poess M, Nambiar R O 2005 Large scale data warehouses on grid: Oracle database 10g and HP proLiant systems *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB'2005) Trondheim Norway* 1055-66

[12] Zhang W S, Cao G H. 2004 *IEEE Transactions on Wireless Communication* **3**(5) 1689-701

[13] Chen H F, Mineno H, Mizuno T 20080 Adaptive data aggregation scheme in clustered wireless sensor networks *Computer Communications* **31**(15) 3579-8

Authors
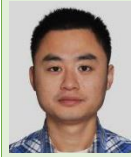
Xingchen Li, born in 1986, Sichuan, China

Current position, grades: Ph.D student of the Department of Information Engineering, Logistical Engineering University of P.L.A.
University studies: B.S. and M.S. degrees from Logistical Engineering University of P.L.A in 2008 and 2011 respectively.
Scientific interest: data processing and decision support system.

Shenglin Li, born in 1964, Sichuan, China

Current position, grades: professor of the Department of Information Engineering, Logistical Engineering University of P.L.A.
University studies: Ph.D. degree in Logistical Engineering University of P.L.A China, in 2008.
Scientific interest: information management engineering and cloud computing.

Heng Zhang, born in 1981, Chongqing, China

Current position, grades: lecturer of the Department of Information Engineering, Logistical Engineering University of P.L.
University studies: B.S. and M.S. degrees from Logistical Engineering University of P.L.A in 2004 and 2008 respectively.
Scientific interest: knowledge engineering, modelling and control of complex systems.

Hui Cai, born in 1982, Chongqing, China

Current position, grades: Ph.D. student of the Department of Information Engineering, Logistical Engineering University of P.L.
University studies: M.S. degrees from Logistical Engineering University of P.L.A in 2010.
Scientific interest: knowledge engineering, modelling and control of complex systems.

Analysis and study on the performance of query based on NOSQL database

Peng Wang, Yan Qi*, Hua-min Yang

School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China

Abstract

With the rise of the Internet web2.0 and the arrival of the era of big data, NOSQL database has already developed rapidly. The traditional relational database in the treatment of high concurrent Web2.0 pure dynamic website and large-scale data encountered bottleneck. This paper introduces the concept of NOSQL database and data storage model. Taking the HBase database as an example, it describes the system structure and data model of HBase, and it demonstrates data query efficiency of the HBase database which is superior to the relational database. In view of the HBase database storage mechanism, the HBase database and MySQL database for performance, which is proceeding comparison, the result shows that NOSQL database performance in processing large data has obvious advantage over the traditional relational database.

Keywords: data model, NOSQL, HBase, MySQL, performance analysis

1 Introduction

In the past 20 years, the traditional relational database has been used widely [1]. With the rapid development of Internet web2.0 and information technology, the size of the data is processed by a computer, which explodes growth. In the face of the emerging Web2.0 applications, traditional relational database cannot meet all kinds of applications, but for the database has higher data concurrent access ability and higher extensibility [2].

With the arrival of the era of big data, the scale of the data to PB or EB order of magnitude, at the same time data type is diversity data type, which mainly includes structured data, non-structured data and semi-structured data. In the face of massive data processing, 80% data is semi-structured and unstructured data. Compared with the structured data (i.e. data, stored in the database, can use the dimension table to the logical expression of the data). It is not convenient to use two-dimensional logical table of the database to represent the data that is called unstructured data, the unstructured data include office documents of all formats, text, images, the subset of XML and HTML of standard generalized markup language, all kinds of report forms, images and audio/video information and so on.

With the rapid development of network and information technology, especially the rapid development of Internet and Intranet technology, the number of non-structured data is increasing. At this time, the mainly limitation is more and more obvious exposed that relational database of management structured data. At this time, NOSQL database is the database oriented application design. At present, industrial and academic circles have the NOSQL database as the new research focus.

2 NOSQL

In today's era of big bank data, academia and industry research and application of large data is continuously deepening. Hence, advantages of NOSQL database is obvious over the traditional relational database.

NOSQL (Not only SQL) refers to the non-relational, distributed, and does not provide the design mode of ACID database. At the same time, NOSQL database does not fixed pattern of data [3]. It is a supplement for the system of relational SQL data.

NOSQL database management system is a different kind of relational database management system, Its data storage format can be loose, it usually does not support the JOIN operation and it is easy to scale horizontally. It can also be called relational database.

2.1 NOSQL DATA STORAGE MODEL

The traditional relational data model often uses the two-dimensional table structure to store and process data [4]. While the NOSQL data storage does not require fixed table structure which is based on the relational model, so there is usually no join operation, which is based on table. While NOSQL has a better performance advantage in data access than the relational database, from the source speaking, this is because the NOSQL data model is based on the requirement of high performance.

According to the data storage model, NOSQL database is divided into many types, including storage, document storage, Key-Value object storage, storage, storage, XML database and so on.

Corresponding author e-mail: 1136161972@qq.com

2.2 RELATIONAL DATABASE AND NOSQL DATABASE

Compared with the relational database, the main advantage of NOSQL data storage management system lies in:

1) Do not make unnecessary complexity. Relational databases provide a wide variety of characteristics and strong consistency, but using the characteristics of relational data in some specific applications, some functionality of relational databases are seldom used. But NOSQL system uses little function to improve performance.

2) High throughput. In terms of throughput, the traditional relational data management system is much lower than some NOSQL database systems.

3) High level extension ability, not high-end hardware cluster. In the aspect of scale, NOSQL database system is able to extend, and do not need too much cost, the cost of scale has a big difference between NOSQL databases and methods of the relational database cluster [5].

4) Object-relational mapping is not going to happen. In terms of large data storage, the data object is stored in a lot of NOSQL database systems, the cost of mutual transformation which is between relational model of database and object model in the program is reduced to zero.

3 HBASE

For large data storage, NOSQL has many mainstream databases. It includes HBase, Redis, CochDB and mongoDB and so on [6]. Below it will take the HBase of NOSQL database and MySQL of relational database as an example, and it explains the difference of performance between HBase and MySQL.

HBase is an open source, distributed, multiple version, the storage model for the column [7]. In addition, loose data is generally stored in HBase. Specifically, the data stored in HBase, which is between Ying Xie (key/value) and relational data [8].

HBase provides a capability similar to Bigtable on Hadoop. HBase is a sub project of the Apache Hadoop project. HBase is different from the relational database in general, it is a suitable for unstructured data storage database. Another difference is HBase, which is based on the column and not based on line model.

3.1 HBASE SYSTEM ARCHITECTURE

HBase server architecture is similar to the HDFS architecture, it follows a simple master/slave server architecture, it is mainly composed of HBase Master Server (HBase Master Server) and HRegionServer Server group [9]. In simple terms, HBase Master Server is responsible for the whole cluster. All of the servers in the HBase are coordinated by ZooKeeper, and ZooKeeper saves the data to HRegionServer mapping. HRegionServer is responsible for storing the large table in the HBase, and

manages read/write data that users send. HBase logic table is defined to be a Region, which is stored in a HRegionServer. Relationship between the HRegionServer and the Region is a one to many relationships.

The internal of HRegionServer manages a series of HRegion objects, each HRegion corresponds to a Region of Table, HRegion is composed of a large number of HStores. Each HStore corresponds to a Column Family in the Table store, we can see that each Column Family which is actually a centralized storage unit, so the best way will put the column of common I/O characteristics into a Column Family, so it is the most efficient way. HStore storage is the core of HBase storage, which is composed of two parts: a part is MemStore, another part is StoreFiles. MemStore is Sorted Memory Buffer, first of all, the user written data will put into the MemStore, when the MemStore is full and it will flush into a StoreFile (the underlying implementation is HFile).When the number of StoreFile file grows to a certain threshold, it will trigger the Compact merge operation, and which multiples StoreFiles into a single StoreFile, the process of the merge will merge and delete data version. Thus it can be seen that : HBase is only increased data, all update and delete operations are performed at the compact during subsequent write operation, which makes the user enter the memory and return immediately, and this will ensure the high performance of HBase I/O. Figure 1 is a HBase architecture diagram.
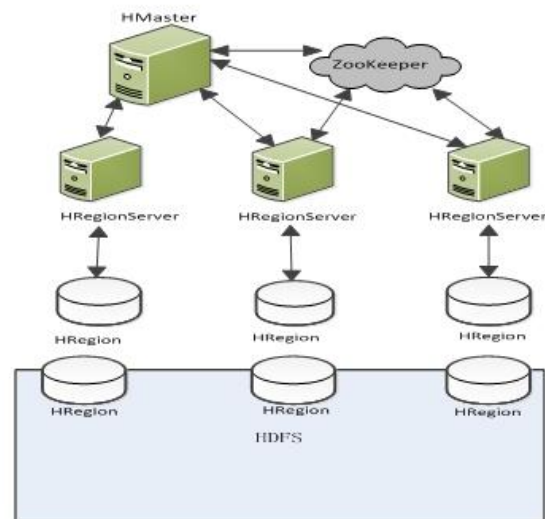
FIGURE 1 HBase architecture diagram

1) Client.

HBase client use RPC mechanism of HBase carries on communication with HMaster and HRegionServer, for the operation of the management, client and HMaster carry on RPC; for the operation of data read and write, the client and HRegionServer carry on RPC.

2) Zookeeper.

Zookeeper Quorum in addition to store-ROOT-address and HMaster address, HRegionServer will also take itself to register to Zookeeper by the Ephemeral, so

that HMaster can be perceived health status of each HRegionServer. In addition, Zookeeper is also to avoid the problems of single point of HMaster.

3) HMaster.

HMaster has no single point problem, HBase can start a number of HMasters, through the Master Election mechanism of Zookeeper to ensure that there is always a Master which is operating, the function of HMaster is mainly responsible for the management of Table and Region:

- manage users on the Table to add, delete, change, check operation;
- manage Load balancing management of HRegionServer, and adjust the Region distribution;
- in Region Split, assign to the new Region;
- when the HRegionServer is shut down, and HMaster is responsible for Regions migration of the failed HRegionServer.

4) HRegionServer.

HRegionServer is mainly responsible for responding to user I/O request, read and write data to HDFS file system, HRegionServer is one of the core modules in HBase.

3.2 DATA MODEL

Big data is special body big, particularly large data sets, the data category, and such data sets cannot use the traditional database tool to grab, management, and its content to its processing [10]. Big data generally refers to the 10 TB (1 TB = 1024 GB) amount of data of scale. But in practical applications, many enterprise users put multiple data sets together, the amount of data has already formed the PB level. In order to solve the limitation of relational database in processing large data theory and implementation, we use the HBase of NOSQL database for data storage.

In the HBase database, data logical structure is also table, but the table structure of the HBase database is different from the relational database. HBase is a sparse long-term storage, multi-dimensional, sort of mapping table. The first row in the table of the HBase is the structure information of table, in the definition of the table. Firstly, to define the column family of the table, the column family is composed of many columns, when the users store the data in the table and each line has a sortable primary key and any number of columns [11].

TABLE 1 Conceptual view of HBase table data

Row key	Timestamp	Column family: C1		Column family: C2	
R1	T3	C1:1	Value1-(1/1)		
R1	T2	C1:2	Value1-(1/2)		
R1	T1			C2:1	Value1-(2/1)

In the HBase table, some columns are blank; the blank columns will not be stored.

1) Conceptual view.

HBase table is conceivable to be a great relationship mapping, it will be able to locate the specific data by row key, row key + a timestamp or row key+ column, Table 1 is the conceptual view of HBase table data.

There is a data in this table: R1, and there are two column families: C1 and C2. Among them, C1 has two data, C2 has a data.

2) Physical view.

From conceptual view of HBase at each table is made up of many lines, but in the physical storage, table is stored in the column, the following Table 2 is the physical view of HBase table data.

TABLE 2 Physical view of HBase table data

Row key	Timestamp	Column family:C1	
R1	T3	C1:1	Value1-(1/1)
R1	T2	C1:2	Value1-(1/2)
Row key	**Timestamp**	**Column family:C2**	
R1	T1	C2:1	Value1-(2/1)

Some columns are blank in the conceptual view, this column will not be stored, actually when request this blank cell, it returns null.

3.3 JAVA API

HBase in the data storage process, the main classes covered include: HBaseAdmin, HTable, Put, Get, Scanner and ResultScanner etc [12]. The following will describe the HBaseAdmin class, HTable class, Scanner class, ResultScanner class, and talk about methods of the above mentioned class in detail.

1) HBaseAdmin.

Relationship:org.apache.hadoop.hbase.client.HBaseAdmin.

Role: provide an interface to manage the information of HBase database table. There is include the methods that it provides: create table, delete the table, list item, make table valid or table invalid, and add or remove column family members of a table and so on.

Table 3 shows the main function of HBaseAdmin class.

TABLE 3 HBaseAdmin class

Function	Describe
addColumn(String tableName, HColumnDescriptor column)	To add a column to an existing table
createTable(HTableDescriptor desc)	Create a table, synchronous operation
deleteTable(byte[] tableName)	Delete an existing table
enableTable(byte[] tableName)	Make the table valid state
disableTable(byte[] tableName)	Make the table invalid state

2) HTable.

Relationship:org.apache.hadoop.hbase.client.HTable.

Role: It can be used to keep communication with HBase directly. This method is not thread safe for update operations.

Table 4 shows the main function of HTable class.

TABLE 4 HTable class

Function	Describe
close()	Release all resources or hang in the update of internal buffer
exists(Get get)	Check whether the Get of the value specified instance exists in HTable column
get(Get get)	Get the Corresponding to some cells access to the values of the specified row
getScanner(byte[] family)	Get scanner instance of the current given column family
getTableDescriptor()	Get HTableDescriptor instance of the current table
getTableName()	Get the table name
put(Put put)	Add the values to table

3) Scanner.

Relationship:org.apache.hadoop.hbase.client.Scan.

Role: According to the conditions to obtain for a batch records.

Table 5 shows the main function of Scanner class.

TABLE 5 Scanner class

Function	Describe
SetCaching()	Set the number of cache, reasonable setting the cache amount, and it can improve the query efficiency
SetBatch()	Specify the maximum number of returned Cell
SetStartRow()	Specify the start line
SetEndRow()	Specify the end of the line (not including the trip)
SetFilter()	Specify a Filter to Filter out unwanted messages

4) ResultScanner.

Relationship: Interface.

Role: The client obtains the value of the interface. Table 6 shows the main function of ResultScanner class.

TABLE 6 ResultScanner class

Function	Describe
Next()	Get the next row values
Close()	Close the scanner and release its allocated resources

Users mainly use the Scanner class and ResultScanner when querying data, in order to illustrate the above Scanner class and ResultScanner class that provide methods, take an example, a logging system has a demand, it is based on an operation to find out all the operations of the whole session. Suppose that we find all the operations of SID = 100000, the code is as the following:

```
Scan scan = new Scan ();
String startkey=DATETIME + "100000"
String endkey= DATETIME +"100000"+"99999"
scan.setStartRow(Bytes.toBytes(startKey));
scan.setStopRow(Bytes.toBytes(endKey));
ResultScanner rs = currentTable.getScanner(scan).
```

By using the Scanner class provides the method of setStartRow and setStopRow, the method of setStartRow and setStopRow limit the search scope, and they effectively improve the query speed of data.

4 Case analysis

The physical store of HBase is based on the column store [13], but the relational database is based on the table structure and mode of storage. The data operation of HBase is different from the data operation of the relational database. For example, HBase only has query, insert, delete, etc, and between the table and table are separated, there is no connection between the table and the table, but relational database has many functions and join operations.

However, structure of HBase table does not need the connection operation, but it can solve the problem that the connection operation solves. The solution is adding a specific keyword in a row [14], you can put all the data of the connection together.

Below we will take a Storage table of the global logistics distribution system as an example to simulate the HBase table design. As is known to all, in the relational database, the storage table structure of the global logistics system is as shown in Table 7 to Table 9. In the three tables, it assumes that the data of a year is about 100 terabytes to 1000 terabytes, in the query of data, the speed will be very slow, so the design of the three tables according to HBase table model, which are designed into two tables, this will effectively improve the query speed of data.

TABLE 7 Storage

S_ID	S_Name	S_*Area*	S_Call	S_Address
Storage ID	Storage Name	Storage Area	Call	Address

TABLE 8 Parts

P_ID	P_Name	P_Rule	P_Unit price	P_Describe
Parts ID	Parts Name	Parts Rule	Unit price	Describe

TABLE 9 Stock

St_uid	St _bid	St _Quantity
Stock uid	Stock bid	Quantity

In the HBase database, data storage mode will be as shown in Table 10 to Table 11.

TABLE 10 Storage table of HBase

Row key	Column Family		Column Family	
<S_ID>	Info	value	parts	value
<S_ID>	Info: S_Name	The name	parts:<P_ID>	<St_Quantity>
<S_ID>	Info: S_Area	The area	parts:<P_ID>	<St_Quantity>
<S_ID>	Info: S_Call	The call	parts:<P_ID>	<St_Quantity>
<S_ID>	Info: S_Address	The address	parts:<P_ID>	<St_Quantity>

TABLE 11 Parts table of HBase

Row key	Column Family		Column Family	
<P_ID>	Info	value	storage	value
<P_ID>	Info: P_Name	The Name	storage:<S_ID>	<St_Quantity>
<P_ID>	Info: P_Rule	The Rule	storage:<S_ID>	<St_Quantity>
<P_ID>	Info: P_Unit price	The Unit price	storage:<S_ID>	<St_Quantity>
<P_ID>	Info: P_Describe	The Describe	storage:<S_ID>	<St_Quantity>

From the above 5 can be seen in the Table, it can be completed operations in a relational database, the operations not only can be finished in HBase, also can have a better performance, the Row key is an index in HBase, Therefore, data query in HBase, query speed is faster than a relational database.

5 Performance test and result analysis

Hardware environment: OS is Ubuntu Server 13.04, HBase version is 0.20.6, Hadoop version is 1.0.1, 10 sets of ordinary PC cluster, on the cluster uses HDFS as a file system [15]. We will compare the performance of NOSQL HBase database with the relational database MySQL, test two for performance difference between 1000000 data query.

Figure 2 to Figure 7 is in the same case, throughput and transactions per second of HBase database and throughput and transactions per second of MySQL database.
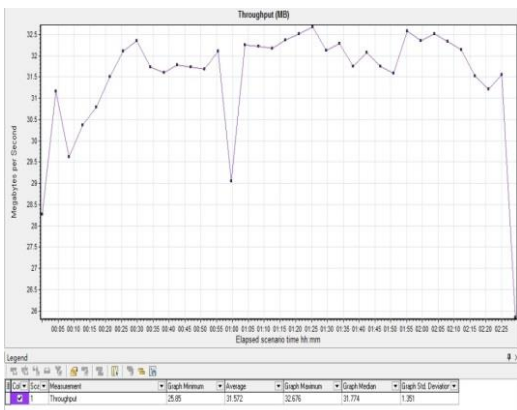
FIGURE 2 Throughput of HBase database
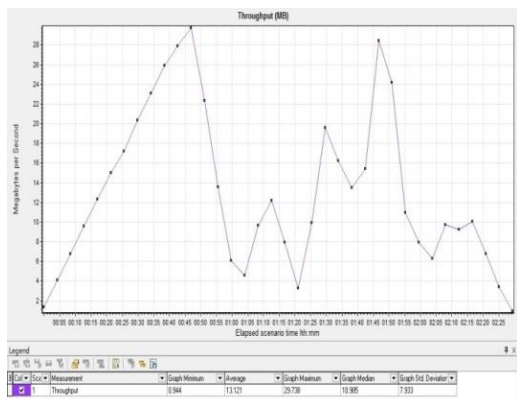
FIGURE 3 Throughput of MySQL database
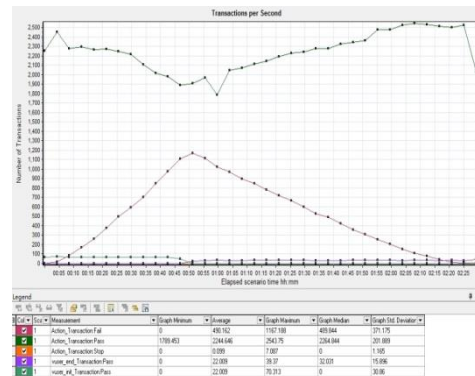
FIGURE 4 Transactions per Second of HBase database
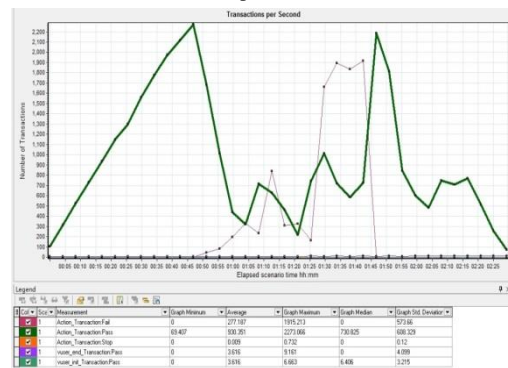
FIGURE 5 Transactions per Second of MySQL database
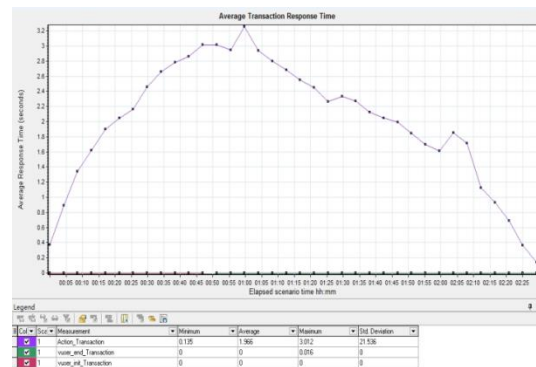
FIGURE 6 Average Transaction Response Time of HBase database
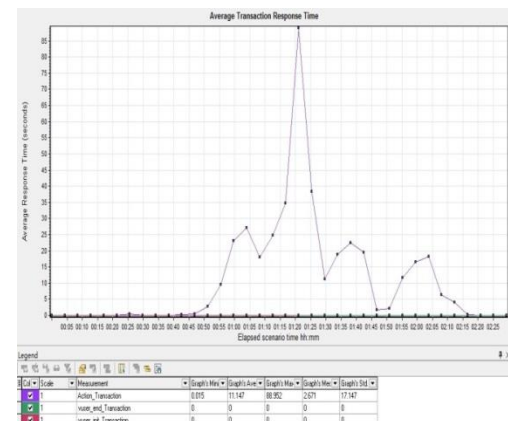
FIGURE 7 Average Transaction Response Time of MySQL database

Seen from the Figure 2 and Figure 3, when operation is the same and the number of concurrent is the maximum, the throughput of HBase is close to the average value, the throughput of MySQL reaches the minimum value in

addition to the starting point and ending point. When the time of test is the same, the average of HBase throughtput is 31.572MB, the average of MySQL throughtput is 13.121MB, the average of HBase throughtput is the average of MySQL throughtput 2.046 times. HBase curve is more stable than MySQL curve under the same condition.

Seen from the Figure 4 and Figure 5, when operation is the same and the number of concurrent is the maximum, HBase and MySQL through the number of transactions reach a low value. HBase in the corresponding numerical and maximum number of the through transactions per second is about 300, while the MySQL in the corresponding numerical and maximum number of the through transactions per second is about 2000.When the time of test is the same, the average of HBase transactions per second is 2244.646, the average of MySQL transactions per second is 930.351, HBase transactions per second is the average of MySQL transactions per second 2.41 times.

Seen from the Figure 6 and Figure7, when operation is the same and the number of concurrent is the maximum, the average transaction response time of the HBase is close to average, average transaction response time of the MySQL reaches the maximum value. When the test time is same, the average value of the HBase average transaction response time is 1.966s, the average value of the MySQL average transaction response time is 11.147s, the average value of HBase average transaction response time average is less than the MySQL average transaction response time.

Table 12 shows the experimental comparison result of HBase database and MySQL database.

From the above data show that: the maximum of the throughput per second in the HBase database is 32.676 MB, the minimum is 25.85MB, the average is 31.572MB. The maximum of the transactions per second in the HBase database is 2543.75, the minimum is 1789.453, the average is 2244.646.

TABLE 12 Experimental comparison result

	HBase			MySQL		
	max	*average*	*min*	*max*	*average*	*min*
Through-put(MB)	32.676	31.572	25.85	29.738	13.121	0.944
Transactions per second	2543.75	2244.646	1789.453	2273.066	930.351	69.407
Average transacttion response time(S)	3.012	1.966	0.135	88.952	11.147	0.015

The maximum of the throughput per second in the MySQL database is 29.738 MB, the minimum is 0.944MB, the average is 13.121MB. The maximum of the transactions per second in the MySQL database is 2273.066, the minimum is 69.407, the average is 930.351.

The maximum of the Average Transaction Response Time in the HBase database is 3.012S, the minimum is 0.135S, the average is 1.966S. The maximum of the Average Transaction Response Time in the MySQL database is 88.952S, the minimum is 0.015S, the average is 11.147S.

The experimental results show that: HBase curve is more stable than MySQL curve under the same condition. Therefore, performance advantage of the HBase database is superior to the MySQL database.

6 Conclusions

This paper introduces the main stream HBase database of NOSQL, and elaborates the data model and related features of the HBase database. It compares the performance of the HBase database with the relational database. The experimental result shows that: the performance of HBase in processing large data is relational database that cannot match advantage. At the same time, NOSQL database in emerging applications also shows a good momentum [16], and NOSQL database has injected new energy into the database of the industry.

References

[1] Hecht R, Jablonski S 2011 NoSQL Evaluation *International Conference on Cloud and Service Computing* 336-41
[2] Jing H, Haihong E, Guan L 2011 Survey on NoSQL database Pervasive Computing and Applications (ICPCA) *6th International Conference* 363-6
[3] Lu J 2013 Big data challenge and NoSQL database technology *Beijing Electronic industry press* (*in Chinese*)
[4] Wang S, Sa S 2006 Introduction to database system *Beijing Higher education press (in Chinese)*
[5] Franke C, Morin S 2011 Distributed Semantic Web Data Mangement in HBase and MySQL Cluster *IEEE 4th International Conference on Cloud Computing* 105-12
[6] Zhu J, Wang J, Zhang J 2014 Based on NoSQL database of research and application technology of data query C*hina railway science (in Chinese)*
[7] Lu J 2012 Hadoop in actual combat *Beijing Mechanical industry press* (*in Chinese*)
[8] HBase:bigtable-like structured storage for hadoop hdfs[EB/OL].http://hadoop.apache.org/hbase 2010.
[9] Borthakur D 2011 The hadoop distributed file system: Architecture and design [EB/OL] http://hadoop.apache.org/hdfs

[10] Zhao G, Yi H 2013 The historical opportunity of the era of large data: industrial transformation and scientific data *Beijing Tsinghua university press (in Chinese)*
[11] Li M, Jia H 2012 Cloud data management of the new data model *Computer engineering and design*
[12] Vora M N 2011 Hadoop-HBase for Large-Scale Data *International Conference on Computer Science and Network Technology* 601-5
[13] Huang J, Ouyang X, Jose J, Wasi-ur-Rahman M, Wang H, Luo M, Subramoni H, Murthy C, Panda D K 2012 High-Performance Design of HBase with RDMA over InfiniBand *Parallel & Distributed Processing Symposium (IPDPS) 2012 IEEE 26th International* 774-85
[14] Zhang Y, Ma Y, Meng X 2011 A kind of efficient space keyword query strategy based on HBase *The wind of science and technology*
[15] Wang Hongyu 2011 Hadoop platform application in cloud computing *Software* **32**(4) 36-9
[16] Quan Zhi, Song Jing-jing 2011 Based on the analysis of NoSQL database technology *With the network information system (in Chinese)*

Authors

Peng Wang, born in June, 1973, Zhuolu County, Hebei Province, China

Current position, grades: associate professor of Changchun University of Science and Technology.
University studies: Master of engineering, doctor of engineering in reading of Changchun University of Science and Technology.
Scientific interest: software engineering and information systems, database and data mining.
Publications: More than 10 papers, 6 Ei retrieval among 10 papers.

Yan Qi, born in November, 1987, Heishan County, Liaoning Province, China

Current position, grades: postgraduate student in reading of Changchun University of Science and Technology.
University studies: database theory and software.

Hua-min Yang, born in November, 1963, Wangqing County, Jilin Province, China

Current position, grades: Dr., Doctoral tutor, professor of Changchun University of Science and Technology.
University studies: virtual reality and digital media, image processing.

Cognitive deviations of information symbols in human-computer interface

Xiaoli Wu[1, 2]*, Yating Ying[1], Feng Zhou[1, 2]

[1]*College of Mechanical and Electrical Engineering, Hohai University, Changzhou, China*

[2]*Institute of Industrial Design, Hohai University, Changzhou, China*

Abstract

In the design of information symbols in human-computer interface, one meaning usually has several design forms. In order to solve cognitive deviations generated during information exchange, in the paper, we tested the recognition rates of common information symbols. The testing results indicated that users under different cultural backgrounds showed significant differences in information symbol cognition. Users prefer to clear and concise information symbols. Users are inclined to understand the surface meaning of information symbols. Through the study of the recognition rates of typical information symbols, we established the perceptual confusion models of information symbols. Based on different models, designers can improve cognitive deviations of existing symbols and design information symbols, which are consistent with user cognition, for reasonable human-computer interaction.

Keywords: human-computer interface, information symbol, cognitive deviation, perceptual confusion

1 Introduction

Human-computer interface (HCI) is the communication medium or means between humans and the computer system, the two-way information exchange platform of various symbols and actions between human and computer. Therefore, various symbols are the important recognition language of information interaction. With the rapid development of interactive multimedia information, the information symbol design is extremely important. Information symbol recognition plays an important role in the interaction between user and interface. Information symbols with poor recognition directly affect interaction means, thus leading to cognitive deviation as well as mistaken understanding and selection [1]. In the paper, we studied cognitive deviation of information symbols, investigated user recognition of information symbols, explored cognitive deviations of various information symbols and proposed the perceptual confusion cognitive model of information symbols.

2 Information symbols in human-computer interface

2.1 SYMBOL INTERACTIVE MODEL

Saussure [2] defined linguistic symbol as the combination of the signifier and the signified the overall formed by signifier and signified. The relation between the signifier and the signified is arbitrary. The combination of the signifier and the signified is stabilized through social conventions to form the social conventions among social members. Initially, the combination of the signifier and the signified are random, indicating that the combination of a

form and a meaning is accidental. During the frequent uses in daily life, the accidental relation is gradually evolved into the stable relation. In the evolution of linguistic symbols, firstly, concepts are defined. Each object is named and corresponding symbols are defined. Secondly, various symbols are connected together to form stable combinations. Thirdly, the relation between a name and a meaning is determined and forms social conventions. In this way, the relation between a symbol and symbol-caused experience is standardized.

In 1954, Wilbur Schramm proposed a famous information interactive model [3] to describe the information communication way that people transmit information and realize meaningful communication through symbols, namely, information source - encoding - symbols - decoding - sink, as shown in Figure 1. In the so-called encoding process, an information sender converts his mood and intention into transmittable symbols according to certain rules. In the so-called decoding process, the symbol receiver gets the meaning of the symbols from the sender through his own life space. A symbol is the intermediary of information interaction. In a certain information interaction system, a symbol has a definite meaning. Symbol combinations follow certain rules, so as to ensure smooth information transmission through symbols. In this way, designer encoding results are consistent with decoding results by receivers. Otherwise, if mutual conventions among people disappear and the transmission process is hampered.

*Corresponding author e-mail: wuxlhhu@163.com
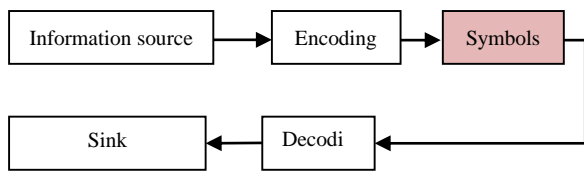
FIGURE 1 An interactive model of information symbols

2.2 RECOGNITION OF INFORMATION SYMBOLS

In the Oxford English-Chinese Dictionary [4], symbols have several basic meanings of text, graphics, images and other visual symbols. In human-machine interface, information symbols belong to graphic symbols, which can convey information. Based on different information types in human-machine interface, information symbols can be divided into text symbols, icon symbols, index symbols, and emblems.

According to symbol design theory, an effectively designed symbol should be not only easily instantly recognized, but also convenient to memory and cognition [5, 6]. Moreover, according to symbol design theory, the familiar concepts are subjected to strangeness treatment and the common and standard concepts are also "creatively corrupted". Then, the new, childish, and vibrant prospect is conveyed through the designed information symbols [7]. In other words, a good visual symbol should be not only easily recognized, but also be reasonably creatively designed. In the creative graphic expression, designers should enhance the recognition of information symbols by the design language.

For example, in the design of the symbol of "entrance", it is necessary to analysis the understanding of the word "enter". Ten entrance symbols from British Railways (BR), London Transport (LT), World Cup (WC), International Union of Railways (UIC), The International Council of Graphic Design Associations (ICOG), the Design of Public Information Symbols of Dreyfuss and Sim are provided in Figure 2. The results of professional fitness analysis and user survey indicate that the most easily recognized symbols are BR and ICOG1. ICOG3 may be mistakenly recognized as "exit" and makes users confused.

As the information conveying way in convey interactive interface, recognition is an important evaluation indicator of the designed symbols [8, 9]. During the graphical symbol creation process, it is necessary to analyse cognitive deviations of information symbols from the perspectives of perception, attention, and memory and then propose the easily instantly recognized symbols with visual impact and without semantic deviation. Therefore, through the study of cognitive deviations of information symbols, the differential analysis of existing information symbols from the perspective of perceptual confusion symbols can contribute to the reasonable design of information symbols.
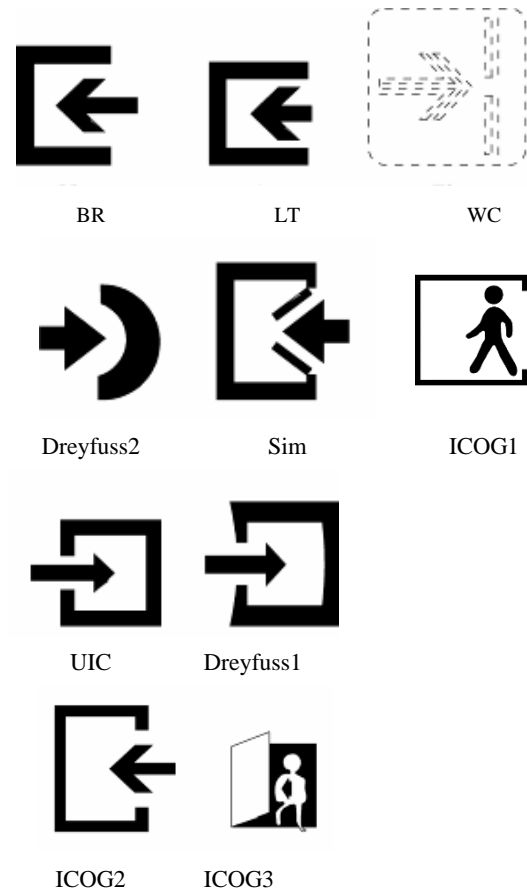
FIGURE 2 Different expression forms of "entrance" symbols

3 Cognitive deviation study of an information symbol

3.1 METHODS

For information symbols represent visual information, simple questionnaire and interview methods can be used to study cognitive deviation. In the questionnaire design, the graphical method is adopted. The interviewees read the symbols and give the symbol meaning. Through the interview, the understandings of various symbols from the interviewees are grasped and personal opinions on information symbols of the interviewees are recorded.

3.2 MATERIALS AND QUESTIONNAIRE DESIGN

Common information symbols are firstly extracted and selected as study objects, so that the interviewees are familiar to the symbols. In the acquired symbols, one concept often has several information symbols or even eight symbols. One concept has different expression forms in different industries and application fields.

TABLE 1 Classification of information symbols

Types	Information symbols
Marks	Positioning map, objective, label, favourite, and grade
Product symbols	Call, photography, video, TV, and phone
Multimedia symbols	Movies, picture, music, volume, voice, equalizer
Outdoor symbols	Train, car, airplane, and dangerous good
Daily symbols	Trash, search, map, compass, clock, lock, unlock, password, rain, and power
Network symbols	Signal, Internet, WIFI, Bluetooth, download, upload, e-mail, information, user, and groups
Document symbols	Pie chart, oscillatory graph, histogram, slide, documents, cut, save, set, tools, and file view
Instruction symbols	Circle and cancel

3.3 PROCEDURE

In the collected information symbols, one concept may have several expression symbols, which lead to different psychological feelings of interviewees. According to age and gender of interviewees, we designed the questionnaire for college students who are sensitive to information symbols of multimedia devices. College students are aged 19-23 years old, including 28 men and 28 women. We obtained 55 questionnaires, including 50 valid questionnaires.

The main survey contents can be divided into three parts: the cognition of different information symbols with the same meaning, understanding and cognition of information symbols, cognitive deviations of similar symbols. After the survey, partial participants are interviewed.

3.4 SURVEY RESULTS

The mathematical and statistical analysis of survey data gives the following results.

3.4.1 The cognition of different information symbols with the same meaning

When users select some common information symbols, such as positioning map, wireless network, and movies, the majority of the users can select the most common symbols and avoid the interference from other symbols, as shown in Table 2. However, when the confusing information symbols coexist, participants can only choose the best answer according to their own experiences. For example, many Chinese are accustomed to considering the symbol of paper clip (⬚) as the bookmark. Although the stripped symbol (⬚) is used as the bookmark in many APPs, Chinese users often improperly select the symbol of paper clip (⬚) as the bookmark because the stripped symbol (⬚) does not meet the Chinese cognitive habits. In addition, if the information symbols are highly similar, participants tend to select concise icons. For example,

FIGURE 3 Information symbols acquired from different interactive interfaces

In the interactive interface of various mobile terminals (mobile telephone and tablet) and different devices (printer, car navigator, camera, and daily electronic devices), investigators extracted common information symbols and collated the acquired symbols. The symbols with single expression form were removed, including "ON" and "OFF". After eliminating colour difference and 3D effect removal, we obtained the testing symbols, as shown in Figure 3.

According to the definition of each information symbol, all information symbols are classified into marks, product symbols, multimedia symbols, outdoor symbols, daily symbols, network symbols, document symbols, and instruction symbols. Representative symbols are selected from the common multimedia devices, classified and organized, as shown in Table 1.

among the symbols with the meaning of "user", most participants select the two most simple and clear options.

TABLE 2 Cognitive deviations of different icons with the same meaning

Types	Information symbols	Participant number of various symbols	Percentages	Meanings	Information symbols	Participant number of various symbols	Percentages
Positioning map		38	76%	Labels		13	26%
		3	6%			21	42%
		4	8%			13	26%
		5	10%			3	6%
Movies		0	0	Users		2	4%
		23	46%			8	16%
		14	28%			4	8%
		7	14%			3	6%
		6	12%			15	30%
						18	36%
Wireless network		33	66%			6	12%
		3	6%			8	16%

3.4.2 Understanding and cognition of information symbols

In the questionnaire, information symbols with recognition difficulty are selected as the questions about understanding and cognition of the meanings. If participants cannot understand the meaning of the information symbol at a glance, it is believed that the symbol has no inherent cognitive thinking.

As shown in Table 3, users tend to understand and select the verbal meaning or to select the superficial meaning of information symbols. On the contrary, the extended meaning of the symbol is seldom considered. For example, the symbol of " ⚡ " in the phone is often used to denote power, but users often understand the symbol of " ⚡ " as "lightning", the superficial meaning of the symbol.

TABLE 3 Cognitive deviations of different icons with the same meaning

	Types	Participant number of various symbols	Percentages
	Files	13	26%
	Place	15	30%
	Store	12	24%
	Classify	7	14%
	Others	3	6%
	Scan	27	54%
	Direction	6	12%
	Measurement	6	12%
	Search	8	16%
	Others	3	6%
	Connection	23	46%
	Circuit	13	26%
	Network	4	8%
	Relation	10	20%
	Power	5	10%
	Lightning	25	50%
	Danger	20	40%
	Thunderstorm	0	0
	Outbox	5	10%
	Upload	35	70%
	Top	5	10%
	Up	5	10%

3.4.3 Cognitive deviations of similar symbols

The analysis of cognitive deviations of different symbols under different meanings indicates that confusing symbols are prone to lead to cognitive deviations among participants. As shown in Table 4, the GPS symbol is "◉", but most participants select "◎". The clipboard symbol is "📋", but most participants select the symbol "📑".

TABLE 4 Cognitive deviations of similar symbols

Documents	35	16	18	15	17
Print	0	6	19	8	17
clipboard	9	22	7	20	8
GPS	24	10	7	11	3
Record	9	9	5	9	36
Position acquisition	10	10	31	26	4
Mobile data	10	42	2	4	23
Cancel	1	3	7	38	12
Refresh	34	1	35	2	9
Tile	10	42	2	4	23
Details	1	3	7	38	12
Medium icon	34	1	35	2	9

3.5 INTERVIEW AND DISCUSSION

Interviews are made after the questionnaire survey is completed. When single participant is interviewed, participants have no definite concept for many information symbols, thus leading to difficulty in the questionnaire survey. The reason of the difficulty can be interpreted in the following three aspects. Firstly, participants have obscure understanding of symbols. Secondly, the recognition of information symbols is poor. Thirdly, participants have limited information symbol cognition and cannot grasp the meanings of information symbols. However, in interviews, participants indicated that easily confused information symbols could not affect their selection deeply. Therefore, information symbol confusion has little effect on cognition. The cognition levels of participants determine the selection of information symbols.

4 Cognitive deviation study of an information symbol

Cognitive deviations of information symbols can be analysed from the cognitive perspective. When a plan is developed to start the implementation according to a certain goal, another goal, plan, or action shows favourable conditions, thus resulting in perceptual confusion. In the design of symbols and command words, distinct symbols should be adopted to avoid visual error, which is named perceptual confusion [10, 11]. Perceptual confusion is one type of cognitive error. Li Leshan [12] divided the users' operation errors into two types: in-attention and over-attention and believed that users' error type could be used to predict the users' intent and to find the thought process of the users. The user errors are divided into four types: slip caused by double capture, forgetting caused by interruption, weakened intentionality, misperception, and over-attention. Norman [13, 14] divided operation errors into the three types: error, slip and mistake. Reason [15, 17] believed that there were 8 basic error types: false sensation, attention failure, memory slip, inaccurate recall, misperception, error judgment, inferential error and unintended actions According to the analysis results of cognitive deviation experiment, we established a cognitive model of perceptual confusion. Perceptual confusion caused by information symbols can be further interpreted as ambiguity, subjective idea, relation misconception, morphological resemblance, complex morphology, and multiple meanings, which lead to recognition difficulty. Information symbols should be designed according to the above reasons of perceptual confusion to reduce cognitive deviation of information symbols (Figure 4).
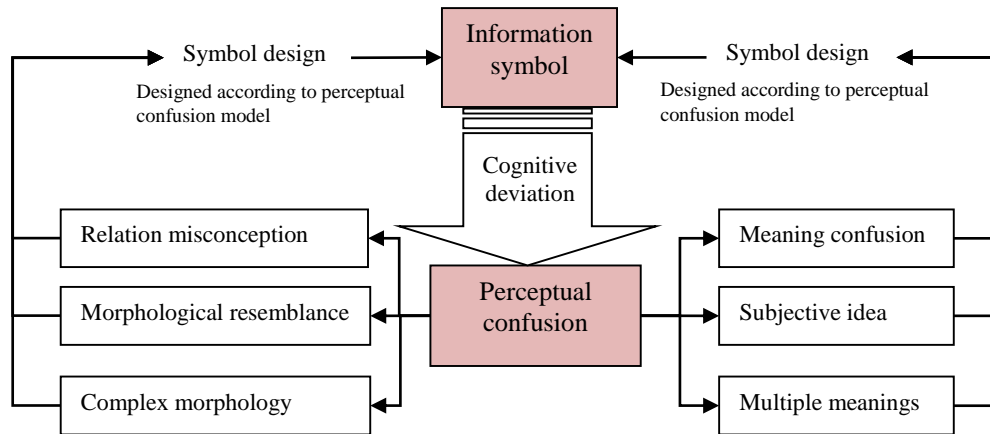
FIGURE 4 Perceptual confusion model of information symbols

5 Conclusions

1) Information symbol design should meet users' habits and cognitive model. Users under different cultural backgrounds show significant difference in cognition of information symbols;

2) Users prefer to clear and concise information symbols, which can avoid excessive cognition burden;

3) Users are prone to grasp the superficial meanings of information symbols and seldom consider its extended meaning;

4) Based on the study of cognitive deviation, we established the perception confusion models of information symbol. According to different perception confusion models, designers should design proper information symbols with less cognitive deviation for smooth human-computer interaction.

Acknowledgement

References

[1] Zheng X 2012 Information-based study of E-commerce website design course *Journal of software* **7**(12) 2794-9 *(in Chinese)*
[2] Saussure F 1980 Cours de linguistique generale *Press The Commercial Press*
[3] Schramm W 1984 Men, Messages and Media:A Look at Human Communication *Press Xinhua Press*
[4] Oxford Advanced Learner's 1988 *Dictionary of Current English with Chinese Translation* Press: The Commercial press
[5] Hawkcs T 1987 Structuralism and semiotics Press Shanghai Translation Press
[6] Gombrich E H 1987 The Senes Of Order *Press Zhejiang Photography Press (in Chinese)*
[7] Ye M 2009 Graphics design: From visual symbol to graphics symbol *Art Research* **2** 101-5 *(in Chinese)*
[8] Annie W Y Ng, Kin Wai Michael Siu, Chetwyn Chan C H 2012 The effects of user factors and symbol referents on public symbol design using the stereotype production method *Applied Ergonomics* **43**(1) 230-8
[9] Piper A K, Boelhouwer E J, Davis J, Holman G T 2008 Using hand drawn images to determine warning symbol design parameters within interactive evolutionary computation software *Proceedings of*

the Human Factors and Ergonomics Society 52nd Human Factors and Ergonomics Society Annual Meeting HFES* 1708-12
[10] Wu X, Huang X, Xu R 2013 An experimental method study of user error classification in human-computer interface *Journal of software* **8**(11) 2890-8
[11] Wu X 2013 Cognitive stratified model on misperception in human-computer interaction *Applied Mechanics and Materials* (364) 849-85
[12] Li L 2004 Human computer interface design *Press Beijing Science Press (in Chinese)*
[13] Norman D A 1988 The Design of Everyday Things *Press New York Doubleday*
[14] Norman D A 1981 Categorisation of action slips 1981 *Psychology Review* **88** 1-15
[15] Reason J 1987 A framework for classifying errors In: New Technology and Human Error *Press In Chichester New technology and Work Wiley*
[16] Reason J 1990 Human Error *Press New York Cambridge University Press*
[17] Reason J 2000 Human error: Models and management *British Medical Journal* **320** 768–70

Authors

Wu Xiaoli, born in July 10, 1980, Yining, China

Current position: PhD candidate at Southeast University, a lecturer at School of Mechanical and Electronics Engineering, HoHai University, China.
University studies: industrial design.
Scientific interest: design-cognition of human-computer interface.
Publications: 4 papers.

Yin Yating, born on September 6, 1991, Suzhou, China

Current position: an interface designer in an Internet company.
University studies: graduated from Hohai University, majored in Digital Media Art.
Scientific interest: the interface design of web games, industrial design.

Zhou Feng, born in May 8, 1972, HuangShan, China

Current position: an associate professor of College of Mechanical and Electrical Engineering, Hohai University, China.
University study: graduated from School of Knowledge Science, Advanced Institute of Science and Technology, Japan
Scientific interest: design creativity and support of creative design process.
Publications: more than 3 papers.

Chinese sentiment analysis for commodity price level fluctuation news comments

Yan Zhao[1*], Suyu Dong[1], Jing Yang[2]

[1]*College of Management, Inner Mongolia University of Technology, Huhhot, China*

[2]*College of Mechanics, Inner Mongolia University of Technology, Huhhot, China*

Abstract

With the rapid development of the Internet technology and news media, people pay more attention on news especially about commodity price fluctuation. Hence, more and more Chinese news comments about commodity price fluctuation appear on Internet. These comments contain all kinds of sentiment. Analysing the sentiment of these comments will make government know more about Netizen emotion on this information and enhance efficiency of management, which has important practical significance. In this paper, we adopt three supervised learning methods (naive Bayes, maximum entropy and support vector machines) to automatically classify user comments as two classes (positive and negative). Through a lot of experiments, we found that machine learning techniques perform quite well in the domain of commodity price fluctuation news comments sentiment classification. Meanwhile, the effects of the feature representations and dimensions for the classification of the three machine learning techniques are analysed and discussed in detail. Experimental results show that maximum entropy classifier is best overall. Frequency is a better method of feature representation, which can use fewer features to get better result.

Keywords: sentiment classification, online reviews, supervised machine learning algorithm

1 Introduction

Nowadays, with the rapid development of the news media, Internet becomes an important part of persons' daily life. The news can be spread with high speed through Internet, which adds people's awareness of paying close attention to online news. Hence, most persons are more likely to comment some news, especially to the news closely linked with daily life, such as news about commodity price fluctuation. Most Netizen comments contain opinion or sentiment, which can impact persons, society and government. As the managers of society, government needs to know the public opinion in time. However, it is time-consuming to browse and analyse all the comments artificially. Hence, an efficient and automatic method of analysis and statistics is necessary.

The technology of sentiment classification aims at analysing texts' opinion automatically. Sentiment classification is widely used in many fields, such as consumption of product, service, social events, vote and so on. Now, in the domain of sentiment classification, most researchers focused on product and service field, few scholars research the classification of new comments. However, news comments contain a lot of public sentiment information about government policies, especially the policies about commodity price fluctuation, which directly impact the person daily life. What's more, people pay more attention to the policies about commodity price

fluctuation. Hence, the analysis and research about these news comments is very necessary.

In this paper, we adopt the methods of machine learning to analyse the sentiment of comments about commodity price fluctuation. We select the three popular machine learning algorithms (naive Bayes, maximum entropy and support vector machines) to classify the sentiment of comments. We also experiment with three feature representation methods (presence, TF, TF-IDF) in the conditions of different feature dimensions. Hence, this paper will analyse and discuss the following problems:

1) Which is the best classifier among SVM, naive Bayes and maximum entropy regarding sentiment classification of commodity price fluctuation news comments?

2) Which method of feature representation for classifiers is the best method regarding sentiment classification of commodity price fluctuation news comments?

3) How the feature dimensions affect the result of classification?

2 Previous Work

With the development of society, Internet has become indispensable part of people's life. Persons pay more and more attention to online news, especially the news closely linked with persons' life. As an important aspect of product and consumption, commodity price fluctuation

---

[*] *Corresponding author* e-mail: yanzhaosky@126.com

has become persons' focus of attention. Nowadays, the clicks of news about commodity price fluctuation rise incessantly, more and more Netizens comment online. In this situation, the large numbers of comments coming from Netizens have had huge impact on society life and public opinion. For the government, the technology of text classification can improve the efficiency of grasping the Netizen comment information.

Sentiment classification aiming at classifying texts according to positive or negative emotion. Sentiment classification has become the research hotpots now, which is a branch of the natural language processing. What's more, the sentiment classification also has large impact on other domains, such as management, sociology, economics and so on. For example, the technology of sentiment classification can help consumers know the information of comments coming from other consumers, enterprise also can grasp the opinion of consumers about the commodities. In addition, government also can grasp the opinion of Netizens as well as the proposals of public. Some researches have applied sentiment classification to some practice fields, such as product, service, financial and so on [1-9], This paper will research the sentiment classification on the news comments of commodity price fluctuation.

In the domain of sentiment analysis, sentiment classification is the most wide research aspect [10-13]. Existing methods of sentiment classification on learning machine can be divided into two classes, which are supervised and unsupervised classification methods. Sentiment classification is one of text classification researches. Hence, existing methods for text classification can be useful for sentiment classification. As a kind of short texts, news comments also can be classified through the methods of traditional text classification. Positive and negative emotions expressed by the emotional words are one of the most important indicators in the problems of sentiment classification. Hence, some researchers used emotional words to classification with unsupervised learning methods [14-17]. [14] used sentiment words, sentiment phrases and fixed syntactic patterns which are used to express sentiment to perform sentiment classification. [15-17] used emotional direction and strength of emotion word, phrase emotional dictionary to sentiment classification, which adopted emotion strength and negative conversion to calculate emotional value to perform classification.

Compared with unsupervised learning methods, supervised machine learning methods are more often used on sentiment classification. Most popular machine learning methods are naive Bayes, maximum entropy and support vector machines. [1] compared naive Bayes, maximum entropy and support vector machines on movies reviews sentiment classification, whose result showed that the feature representation method of presence is better than TF, naive Bayes is better than SVM with TF, SVM with presence is better than naive Bayes, maximum entropy is worst among the three classifiers. [18] compared SVM,

naive Bayes, N-gram semantic model and found that SVM and N-gram semantic model are better than naive Bayes on classification of travel reviews. [19] proved that SVM is not better than naive Bayes for literature reviews classification at all times. [20] showed that naive Bayes is a better classifier than SVM. The examples listed above show that no machine learning algorithm can always maintain the best effect for all kind of texts. Hence, It is necessary to compare naive Bayes, maximum entropy and SVM to find the best methods in commodity price fluctuation news comments classification. In addition, other factors such as feature representation method and feature dimensions are also worth researching.

3 Methodologies

Figure 1 shows the fundamental steps of text classification using supervised learning machine methods. As is known from fundamental theory, the supervised learning machine methods can train labelled texts to get models, which can be used for predicting new unlabelled texts.

The following part of this section will introduce the theory of every section, including feature selection methods, feature representation methods, and classifiers. In this paper, each document is represented as a vector with feature weights. Let $\{t_1, t_2, ..., t_m\}$ be a predefined set of m features that can appear in a document. Let n_i be the number of times that t_i occurs in document d. Let w_i is the feature weight in a document. Each document d is represented by the document vector $d = \{w_1, w_2, ..., w_m\}$.
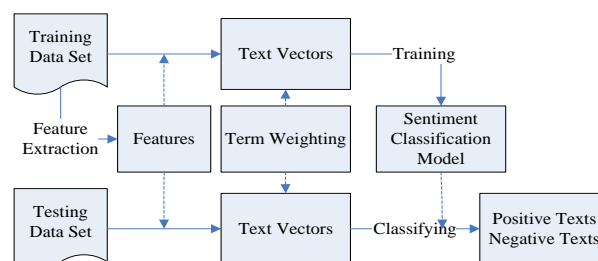
FIGURE 1 Fundamental theory of text classification

3.1 FEATURE SELECTION

Large numbers of features will be produced through feature identification. If all the features are used for classifier, the result and efficiency of classification will be reduced. Hence, feature selection is essential for classification. This paper adopts document frequency feature selection method, whose another name is DF. Generally, the DF firstly counts the number of every feature (DF value) appearing in all texts (comments), and then get the proper features according to the DF value. In our experiment, all the features are sorted according to the DF value. We select top n features to do experiments, The value of n is from 50 to 2950. The 30 groups of features will be selected, the distance of numbers is 100 among each group.

3.2 FEATURE REPRESENTATION

Feature weighting denotes the importance of a feature of a text, namely the distinguish ability of a feature of the text. Feature weight is calculated through the statistical information of text. This paper will compare three different calculation methods of feature weight for classification, they are Boolean, frequency and TF-IDF.

3.2.1 Presence representation

Boolean is based on the feature whether or not appears in the text. When the feature appears in the text, the value is 1, otherwise the value is 0. This method is replaced by other more accurate methods because the Boolean value cannot reflect the importance of feature in text. However, this method can also obtain a good effect under some circumstances. For instance, [1] performed the sentiment classification of movie reviews, which showed that SVM with unigrams model combining Boolean feature representation is better than the other methods of feature representation in the same circumstances.

3.2.2 TF

The method of TF uses the times of feature appearance in the text to represent the text. When we use frequency as the calculation of feature weight, the distinguish ability of low-frequency features will be ignored. However, some low-frequency features may have a greater ability to distinguish text than high-frequency features. [21] used naive Bayes with TF feature representation method performed the standard topic-based categorization and the accuracy is highest. [19] found that frequency is the better feature representation method for SVM and naive Bayes in the novel text sentiment classification experiment.

3.2.3 TF-IDF

TF-IDF is the most widely used feature weight calculation method for the text classification. It is based on the idea: if one feature has high-frequency, and rarely appears in other text, then the feature has a good ability to distinguish. Although it is ideas and structure of statistics are very simple, but its performance is very good. The $TF-IDF$ value of a certain feature is calculated by the following equation:

$$w_{ij} = tf_{ij} \times \log \frac{N}{n_i},\qquad(1)$$

where w_{ij} indicates the weight of feature t_i in document d_j. tf_{ij} indicates the frequency of feature t_i in document d_j. n_i indicates the number of document which contains feature t_i. N is the number of all documents.

3.3 CLASSIFIERS

3.3.1 Naive Bayes classifier

Naive Bayes classifier is a kind of simple classifier, but widely used in text classification. According to the Bayes formula, the probability of document d belongs to C_i is calculated by the following equation:

$$P(C_i|d) = \frac{P(d|C_i) * P(C_i)}{P(d)},\qquad(2)$$

where $P(C_i)$ indicates the probability of a document belonging to C_i. In this paper, we used Naive Bayes classifier with weight. The equation is following:

$$P_{NB}(C_i|d) = \frac{P(C_i)\left(\prod_{t_i \in V} P(t_i|C_i)^{W(t_i,d)}\right)}{\sum_j \left[P(C_j)\prod_{t_i \in V} P(t_i|C_j)^{W(t_i,d)}\right]},\qquad(3)$$

where feature t_i is independent of document d, $W(t_i,d)$ indicates the weights of feature t_i in document d. $P(t_i|C_i)$ indicates the Laplacean probability estimation value of conditional probability of documents belonging to C_i if it contains feature t_i. $P(t_i|C_i)$ is calculated by the following equation:

$$P(t_i|C_i) = \frac{1+W(t_i,C_i)}{|V| + \sum_j W(t_j,C_i)},\qquad(4)$$

where $W(t_i,C_i)$ indicates the number of documents containing features t_i and belonging to C_i. $|V|$ is the size of $\{t_1,t_2,...,t_m\}$, which are all features coming from all documents.

Naive Bayes classifier is based on the assumption of independence conditions, using the joint probability between features and categories to estimate the probability of categories given a document. Although it's assumption conditions is very restrictive and difficult to meet in real-world, it still performed well in text classification [20,22]. [23] showed that naive Bayes can well complete two opposite case data classification, completely independent features classification or functionally dependent features classification.

3.3.2 Maximum entropy classifier

Maximum entropy classifier (ME) is based on maximum entropy model, [24] was the first application of maximum entropy models in the natural language processing; [25] improved maximum entropy model. [26] found that ME is better classifier than Naive Bayes classifier on text classification. Its basic idea is that it does not make any

hypothesis and remain maximum entropy for the unknown information, this is an advantage for maximum entropy compared with Naive Bayes. Maximum entropy model must satisfy the constraint of known information and the principle of maximum entropy. Hence maximum entropy model is got through solving a optimization problem with constraints. The classical algorithm to solve this problem is lagrange multiplier method. In this paper, we give the conclusion directly. The result is following:

$$p^*\left(C_i|t_i\right) = \frac{1}{\sum\limits_{C_i} \exp\left(\sum\limits_i \lambda_i f(t_i, C_i)\right)} \exp\left(\sum\limits_i \lambda_i f(t_i, C_i)\right), (5)$$

where $p*$ indicates a predictive model for classification; V indicates the feature vectors; C_i indicates the type which the document belongs to. λ_i indicates the feature weight of feature vectors containing many feature t_i. $f(t_i, C_i)$ is an indicator function.

3.3.3 SVM

Support vector machine (SVM) is generally considered as the best classifier for traditional text classification [27], it is usually better than naive Bayes and maximum entropy. Naive Bayes and maximum entropy are based on probability model, support vector machine (SVM) classifier is got by solving the optimal hyperplane represented by vector $\vec{W}$. Hyperplane is shown in Figure 2, this hyperplane is used to accomplish classification which can ensure maximum separation between a certain amount of data from the training set and hyperplane. Solving the maximum margin hyperplane eventually is converted into solving a convex quadratic programming problem.
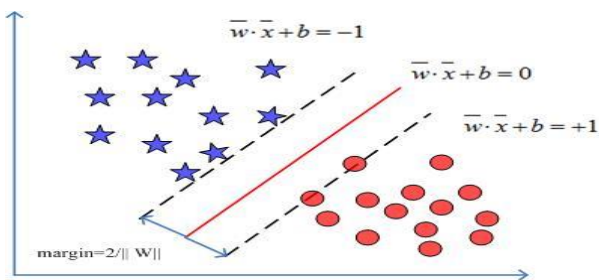
FIGURE 2 The optimal hyperplane

Generally, it translates the above problem into the constrained optimization problem of dual variables through Lagrange Duality. The solution can be written as:

$$\vec{W} = \sum_{i=1}^{n} \alpha_i C_i \vec{d}_i, \tag{6}$$

C_i is the correct category for document $\vec{d}_i$. α_i are support vector and greater than zero.

Moreover, kernel function can be used for linear inseparable problems for SVM to convert low dimensional space nonlinear problem to high dimension space linear problem. Mapping of kernel function can be a good control of the computational complexity of nonlinear expansion and can avoid the curse of dimensionality. There are many kernel functions: linear kernel, Gaussian kernel function, radial basis function and so on. In this paper, we used linear kernel function and optimize the parameter of SVM model, which will be used for following experiments.

4 Experiments

4.1 EXPERIMENT DATA

This paper research the sentiment classification of online comments to news of commodity price fluctuation, therefore, we created a corpus by retrieving reviews of news of commodity price fluctuation from each big Chinese news website (URL: http://news.sina.com.cn; http://news.sohu.com; http://news.qq.com; http://news.163.com; http://www.people.com.cn). The data we used for experiment were downloaded by a crawler and randomly selected 3566 comments from all comments.

In this paper, we focus on classifying comments as positive or negative. However, there is no label about the polarity of sentiment for comments. Thus, three students were trained to annotate these comments. In the whole process of the annotation, non-commodity-price-fluctuation news comments were excluded before annotating the polarity of sentiment. Comments were annotated polarity according to the unified label; the values were 0, 1, 2. Thereinto, 0 represents the negative comments, 1 represents the positive comments, 2 represents the comments which its polarity cannot be judged. We found that there was inconsistent between students when they annotated comments. We weeded out the comments if:

1) it is annotated differently by three students;
2) two students' judgment annotates it with 2.

Finally, 1500 negative comments and 1500 positive comments were randomly chosen to establish the corpus.

4.2 EVALUATION METHOD

In this paper, the results of sentiment classification are evaluated by three indexes that are frequently used in text classification: Accuracy, Precision and Recall. Accuracy is used to justify the overall performance of sentiment classification. Precision and recall are used to evaluate the performance the negative and positive classification. These indexes can be calculated according to Table 1.

TABLE 1 Results of experiments

	Classified positive comments	**Classified negative comments**
labelled positive comments	a	b
labelled negative comments	c	d

The calculation equations are the following, respectively:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \, , \tag{7}$$

$$\text{Precision}(pos) = \frac{a}{a+b} \, , \tag{8}$$

$$\text{Precision}(neg) = \frac{d}{c+d} \, , \tag{9}$$

$$\text{Recall}(pos) = \frac{a}{a+c} \, , \tag{10}$$

$$\text{Recall}(neg) = \frac{d}{b+d} \, . \tag{11}$$

5 Experiment result and discussion

We adopt 3-fold cross-validation to do experiment. We adopt our own implementation for text pre-processing, NLPIR toolkit is used for Chinese text segmentation, McCallum's Mallet toolkit [28] implementation of naive Bayes classifier and maximum entropy classifier and Chang's LIBSVM [29] implementation of a Support Vector Machine classifier are used for classification. We ran each classifier with various feature representations and different number of features to experiment.

5.1 EXPERIMENT RESULTS OF NB CLASSIFIER

The performances of sentiment classification of NB with various feature representations and different feature sizes are showed in Figures 3-5.
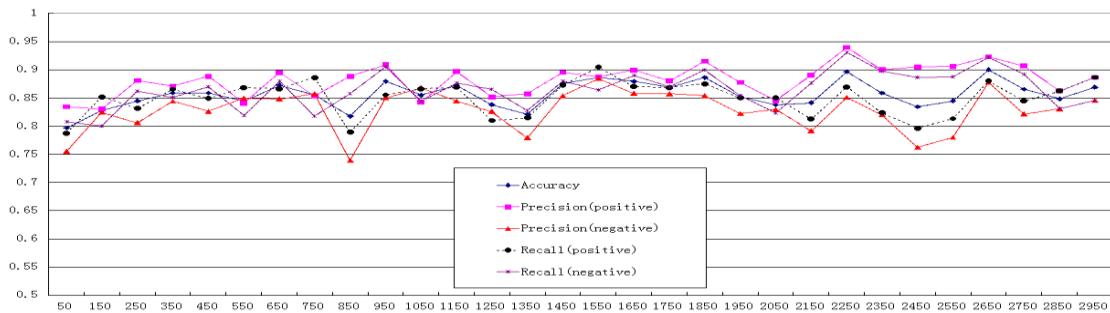
FIGURE 3 NB with presence representation under different feature dimensions
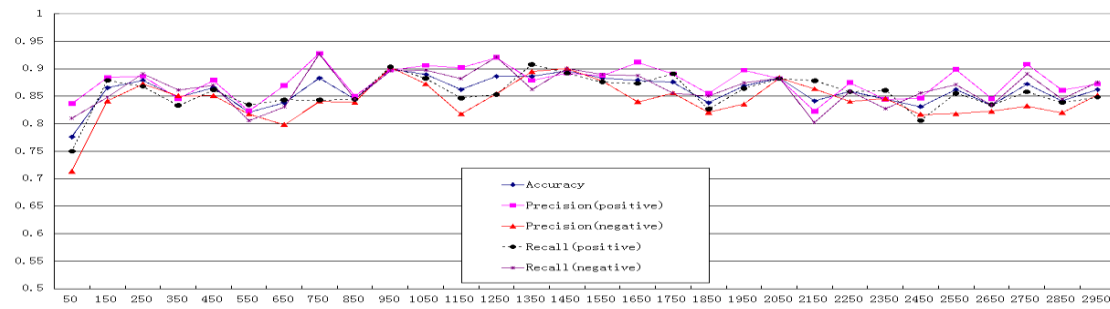
FIGURE 4 NB with TF representation under different feature dimensions
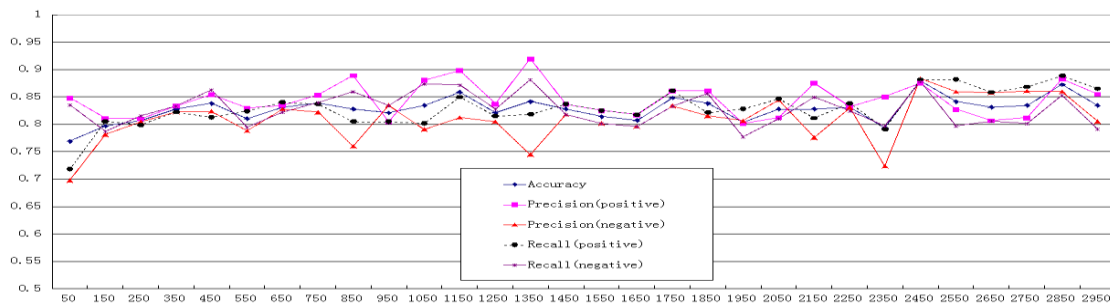
FIGURE 5 NB with TF-IDF representation under different feature dimensions

171

Figure 3 indicates the result of NB by accounting for feature presence, the most results of average accuracy are between 85% and 90%, the average accuracy peaks at 90% with 2650 features and the lowest average accuracy is 79.66% with 50 features. Figure 4 displays the result of NB with TF feature weight calculation method, the average accuracies distribute mainly between 85% and 90%, the top of average accuracy is 90% with 950 features, and the minimum average accuracy is 77.59% with 50 features. Figure 5 shows the results of NB with TF-IDF feature representation, the most average accuracies are between 80% and 85%, the peak of average accuracy is 87.93% with 2450 features and the minimum average accuracy is 76.90% with 50 features.

5.2 EXPERIMENT RESULTS OF ME CLASSIFIER

The performances of sentiment classification of ME with various feature representations and different feature dimensions are showed in Figures 6-8. Figure 6 indicates the result of ME by accounting for feature presence, the average accuracy peaks at 91.03% with 1250 (2350) features and the lowest average accuracy is 77.24% with 50 features. Figure 7 displays the result of ME with TF feature weight calculation method, the top of average accuracy is 91.38% with 2250 features, the minimum average accuracy is 74.83% with 50 features. Figure 8 shows the results of ME with TF-IDF feature representation, the peak of average accuracy is 91.72% with 1750 features and the minimum average accuracy is 80.00% with 50 features. As Figure 6-8 show, 90% results of the average accuracy of ME with three feature representations are above 85%.
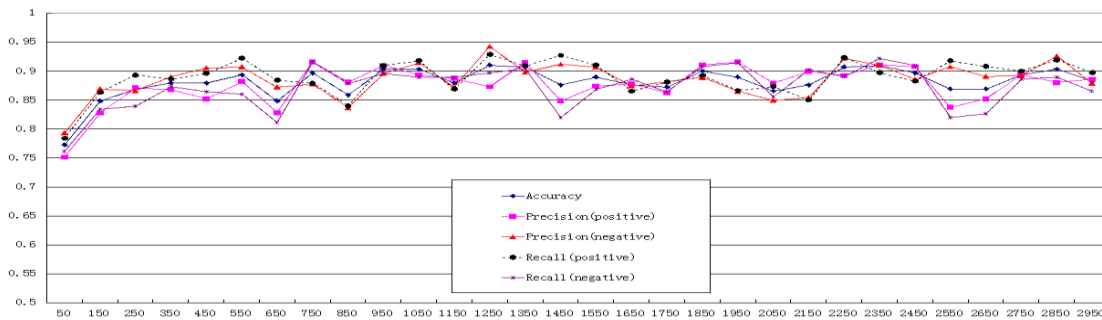
FIGURE 6 ME with presence representation under different feature dimensions
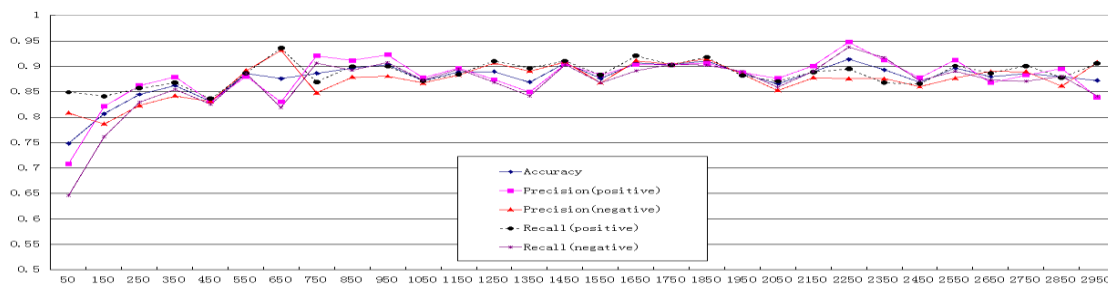
FIGURE 7 ME with TF representation under different feature dimensions
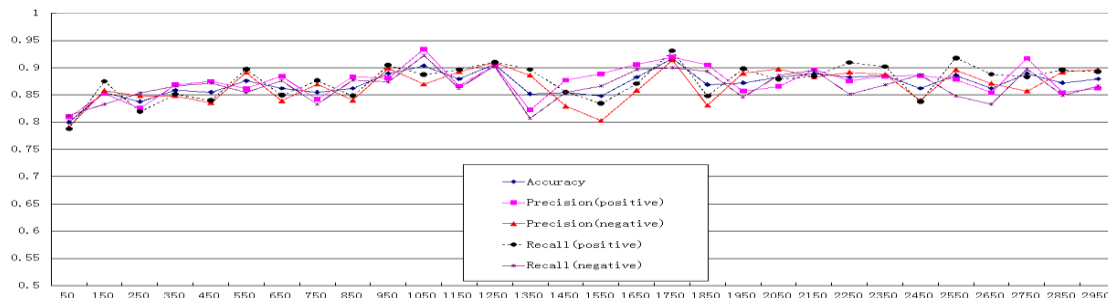
FIGURE 8 ME with TF-IDF representation under different feature dimensions

5.3 EXPERIMENT RESULTS OF SVM CLASSIFIER

The performances of sentiment classification of SVM with various feature representations and different feature sizes are showed in Figures 9-11. Figure 9 indicates the result of SVM by accounting for feature presence, the range of average accuracy is wide, the average accuracy peaks at 89.35% with 1850 features and the lowest average accuracy is 56.36% with 2250 features. Figure 10 displays

the result of SVM with TF feature weight calculation method, more than 90% results of the average accuracy are above 85%, the top of average accuracy is 87.97% with 350 features, the minimum average accuracy is 80.76% with 50 features. Figure 11 shows the results of SVM with TF-IDF feature representation, the most average accuracies are around 85%, the peak of average accuracy is 87.63% with 2550 features and the minimum average accuracy is 80.07% with 50 features.
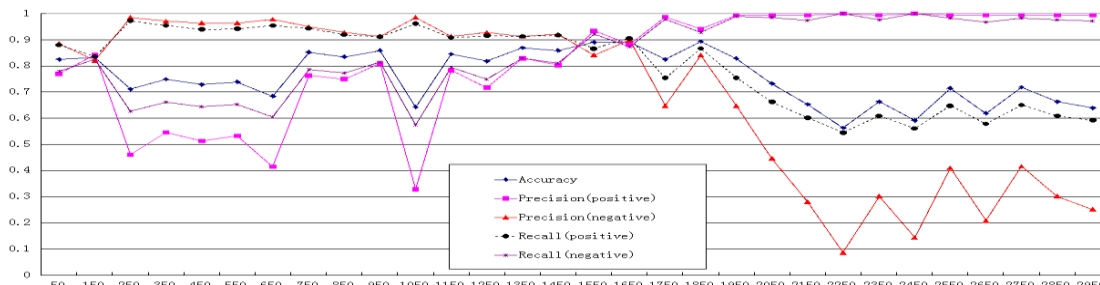
FIGURE 9 SVM with presence representation under different feature dimensions
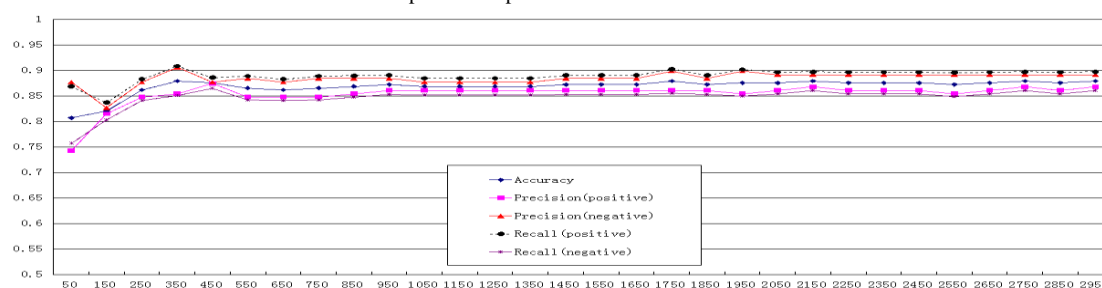
FIGURE 10 SVM with TF representation under different feature dimensions
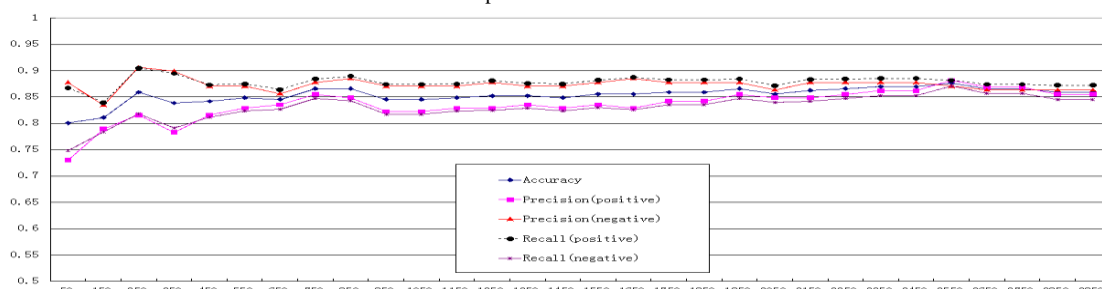
FIGURE 11 SVM with TF-IDF representation under different feature dimensions

5.4 COMPARISON AND ANALYSIS OF EXPERIMENT RESULTS

5.4.1 NB vs ME vs SVM

As Figures 3-11 show, the top of average accuracy will be varied according to different classifiers with different calculation methods of feature weight. Table 2 shows the statistical result of highest average accuracy for every classifier with three feature representations, ME is the best classifier for sentiment classification of online comments to news of commodity price fluctuation. The tops of average accuracy of ME with three feature representations are all above 91.00%, thereinto, the best average accuracy is 91.72% accounting for TF-IDF. ME slightly outperform NB. The highest average accuracy is 90% of NB with presence and TF feature weight calculation methods, but the number of features for NB with presence and TF feature representations are 950 and 2650, respectively.

TABLE 2 The top of average accuracy under three classifiers with three

	Presence	**TF**	**TF-IDF**
NB	90.00%	90.00%	87.93%
ME	91.03%	91.38%	91.72%
SVM	89.35%	87.97%	87.63%

Table 2 indicates that the minimum value of the top of average accuracies of three classifiers with three feature representations is 87.63%, which is achieved by SVM with TF-TDF. Table 3 shows the result of the number of features, if and only if the features are needed for three classifiers with different feature representations when their average accuracies are 87.63%. We find that ME achieve accuracy around 87.63% using few features than NB and SVM, especially ME with presence and TF-IDF feature weight calculation methods. NB requires fewer features than SVM with three feature representations.

TABLE 3 The feature dimensions when average accuracy achieve 87.63% in first time

	Presence	TF	TF-IDF
NB	950	250	2450
ME	350	550	950
SVM	1550	350	2550

The negative precision and positive precision of three classifiers with different feature weight calculation methods is varied. Table 4 displays the sum of absolute of D-value from negative precision and positive precision of three classifiers. For NB, with three feature representations, the positive precision is higher than negative precision. SVM achieves better negative precision. However, the sum of SVM with presence is more than others. Figure 9 shows the difference between negative precision and positive precision of SVM with presence is larger. The gap between positive precision and negative precision of ME is smallest.

TABLE 4 The sum of absolute of D-value from negative precision and positive precision

	Presence	TF	TF-IDF
NB	1.676	1.166	1.466
ME	0.932	0.951	1.002
SVM	12.197	0.909	1.165

The stability of three classifiers with three feature representations is different. Table 5 shows the mean values and variances of average accuracy. The variances of SVM with TF and TF-IDF is less than other variances, but the variance of SVM with presence is largest. Although ME outperform SVM and NB, its variance is larger. The variances of NB are smaller than the variances of ME. Thus, the result demonstrates a descending order of the stability of three classifiers as SVM (TF, TF-IDF)>NB>ME. The performance of SVM with presence is most unstable. In practice application, if you pay attention to the stability of classifiers, you can select SVM and TF to achieve higher average accuracy to sentiment classification of online news comments of commodity price fluctuation.

TABLE 5 The mean values and variances of average accuracy

	mean values of average accuracy (%)			variances of average accuracy		
	Presence	TF	TF-IDF	Presence	TF	TF-IDF
NB	85.62	86.06	82.77	0.0240	0.0265	0.0224
ME	88.0	87.66	87.10	0.0270	0.0336	0.0228
SVM	77.58	86.93	85.38	0.0993	0.0158	0.0160

5.4.2 Presence vs TF vs TF-IDF

For presence, TF and TF-IDF, combining with three classifiers to sentiment classification of online comments to news of commodity price fluctuation, the best average accuracy is achieved by TF-IDF and ME. However, different classifiers are suitable for different feature representations. NB with TF can use small size of features

to achieve higher average accuracy. SVM with presence achieves the best average accuracy.

Table 3 shows that, three classifiers with different feature representations achieve the same average accuracy, they use a fewer features when they adopt TF feature weight calculation method than presence and TF-IDF. As Table 2 and Table 5 show, compared with other feature representations, TF has the superiority on the aspect of top of average accuracy and mean value of average accuracy. Table 4 displays that, the gap between negative precision and positive precision of TF is smallest. In this paper, TF is the best feature weight calculation method and its compatibility is best.

5.4.3 The number of features

As is shown above, the results of sentiment classification of online comments to news of commodity price fluctuation, except the accuracy of SVM with presence, which can use 50 features to achieve 82.47% accuracy, the average accuracy of three classifiers with different feature representations with low feature dimension is low. We find that the accuracies are low when classifiers with few features, but with the number of features increasing, the accuracies of three classifiers reach their peaks and then decline or fluctuate. For instance, the average accuracy of SVM with presence peaks 89.35% with 1850 features and then declines, the average accuracy of ME with presence peaks 91.03% with 1250 features and then fluctuates. This proved that, the effects of classifiers are influenced by feature dimensions when few features are used for classification, because the helpful features are not included. Moreover, after the accuracies reach their peaks, the effect of feature dimensions for classifiers is small, and the average accuracy can be improved through perfecting classifiers, feature weight calculation method and feature extraction method.

As Figures 2-11 show, according with the formula of recall and precision, the trend of positive recall and negative precision is same, the trend of negative recall and positive precision is same. What's more, the experiment result of recall and precision also proved the veracity of classifiers.

5.4.4 Model analysis

Overall, the maximum entropy classifier has outstanding performance for commodity price fluctuation news comments classification compared to naive Bayes and SVM. The deep reasons are related to the maximum entropy theory and character of these comments. The commodity price fluctuation news comments classification has the peculiarity that a lot of comments express the subjective opinion through the objective words, which enhance the difficulty of classification. However, the maximum entropy theory is based on that it does not make any hypothesis and remain maximum entropy for the unknown information. Hence, maximum

Zhao Yan, Dong Suyu, Yang Jing

entropy classifier has strong robustness for our text data. Compared to two other classifiers, the theory of naive Bayes classifier is based on assumption that the features are independent with each other. In fact, the dependent relationship is existing between features. Hence, the classification result of naive Bayes is little worse than maximum entropy. For SVM classifier, the presence feature representation method has instability. When the feature dimensions is large enough, the classification accuracy reduce rapidly. However, the TF and TF-IDF feature representation methods for SVM have strong stability. The reason is that SVM classifier with presence feature representation will result in data sparse problem when the feature dimensions is very high, which make the accuracy reduce.

6 Conclusions

Since 2000, sentiment analysis has become a very active research area in linguistics and natural language processing. Although there are many researches about sentiment classification, most research is about product and service and little research had been done about sentiment classification of news comments. In this paper, we focus on the sentiment classification of online comments to news of commodity price fluctuation. We analyze the characteristics of Naive Bayes model, maximum entropy model and SVM model.

From the performance, we find that machine learning techniques perform quite well in the domain of sentiment classification of online comments to news of commodity price fluctuation. Comparing NB and SVM, ME is the most effective and efficient. The top accuracy of three classifiers is 92.72%, which achieves by ME with TF-IDF. However, SVM is the most stable classifier. With different feature representations, the accuracies of three classifiers reach their peaks. Considering feature dimensions simultaneously, TF is the best feature weight calculation method. The results of experiment proved that the effects of classifiers are affected if the number of features is small. With the increasing of dimensions of features, the influence of features dimensions is reducing. In the practical application, the dimensions of features should be chosen properly. Only in this way, can the result of classification will be efficient and accurate.

The value of this research is big, it can be useful for different areas, such as governments, business and so on. On the basis of this research, future research will be extended to the sentiment analysis on cloud platforms to solve all kinds of big data problems.

Acknowledgments

References

[1] Pang B, Lee L, Vaithyanathan S 2002 Thumbs up?: sentiment classification using machine learning techniques *Proceedings of the ACL-02 conference on Empirical methods in natural language processing Association for Computational Linguistics* **10** 79-86

[2] Turney P D, Littman M L 2003 Measuring praise and criticism: Inference of semantic orientation from association *ACM Transactions on Information Systems (TOIS)* **21**(4) 315-46

[3] Mullen T, Collier N 2004 Sentiment Analysis using Support Vector Machines with Diverse Information Sources *EMNLP 2004* **4** 412-8

[4] Kennedy A, Inkpen D 20006 Sentiment classification of movie reviews using contextual valence shifters *Computational Intelligence* **22**(2) 110-25

[5] Tang H, Tan S, Cheng X Q 2007 Research on sentiment classification of Chinese reviews based on supervised machine learning techniques *Journal of Chinese information processing* **21**(6) 88-94

[6] Devitt A, Ahmad K 2007 Sentiment polarity identification in financial news: A cohesion-based approach *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* 2007 984-91

[7] Tian F, Gao P, Li L, Zhang W, Liang H Qian Y Zhao R 2014 Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems *Knowledge-Based Systems* **55** 148-64

[8] Li W, Xu H 2014 Text-based emotion classification using emotion cause extraction *Expert Systems with Applications* **41**(4) 1742-9

[9] Guan W, Gao H, Yang M, Li Y, Ma H, Qian W, Cao Z, Yang X 2014 Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events *Physica A: Statistical Mechanics and its Applications* **395** 340-351

[10] Pang B, Lee L 2008 Opinion mining and sentiment analysis *Foundations and trends in information retrieval* **2**(1-2) 1-135

[11] Haney C 2014 Sentiment Analysis: Providing Categorical Insight into Unstructured Textual Data *Social Media, Sociality, and Survey Research* 35-59

[12] Balahur A, Mihalcea R, Montoyo A 2014 Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications *Computer Speech & Language* **28**(1) 1-6

[13] Balahur A, Turchi M Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis *Computer Speech & Language* **28**(1) 56-75

[14] Turney P D 2002 Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews *Proceedings of the 40th Annual Meeting on Association For Computational Linguistics* 417-24

[15] Hu M, Liu B 2004 Mining and summarizing customer reviews *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining ACM 2004* 168-77

[16] Ding X, Liu B, Yu P S 2008 A holistic lexicon-based approach to opinion mining *Proceedings of the 2008 International Conference on Web Search and Data Mining ACM 2008* 231-40

[17] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M 2011 Lexicon-based methods for sentiment analysis *Computational linguistics* **37**(2) 267-307

[18] Ye Q, Zhang Z, Law R 2009 Sentiment classification of online reviews to travel destinations by supervised machine learning approaches *Expert Systems with Applications* **36**(3) 6527-35

[19] Yu B 2008 An evaluation of text classification methods for literary study *Literary and Linguistic Computing* **23**(3) 327-43

[20] Zhang Z, Ye Q, Zhang Z, Li Y 2001 Sentiment classification of Internet restaurant reviews written in Cantonese *Expert Systems with Applications* **38**(6) 7674-82

[21] McCallum A, Nigam K A 1998 comparison of event models for naive bayes text classification *AAAI-98 workshop on learning for text categorization* **752** 41-8

[22] Lewis D D 1998 Naive (Bayes) at forty: The independence assumption in information retrieval *Machine learning: ECML-98* Springer Berlin Heidelberg 4-15

[23] Rish I 2001 An empirical study of the naive Bayes classifier *IJCAI 2001 workshop on empirical methods in artificial intelligence* **3**(22) 41-6

[24] Berger A L, Pietra V J D, Pietra S A D 1996 A maximum entropy approach to natural language processing *Computational linguistics* **22**(1) 39-71

[25] Chen S F, Rosenfeld R 2000 A survey of smoothing techniques for ME models *IEEE Transactions on Speech and Audio Processing* **8**(1) 37-50

[26] Nigam K, Lafferty J, McCallum A 1999 Using maximum entropy for text classification *IJCAI-99 workshop on machine learning for information filtering* **1**: 61-7

[27] Joachims T 1998 Text categorization with support vector machines: Learning with many relevant features *Springer Berlin Heidelberg*

[28] McCallum, Andrew Kachites 2002 MALLET: *A Machine Learning for Language Toolkit* http://mallet.cs.umass.edu

[29] Chih-Chung Chang and Chih-Jen Lin 2011 LIBSVM : *a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Authors

Yan Zhao, born in July, 1982, Tong Liao, Inner Mongolia, China

Current position, grades: Associate Professor at the College of Management, Inner Mongolia University of Technology.
University studies: Doctor's degree in Economics from Beijing Central University of Finance and Economics in June 2011.
Scientific interests: internet public opinion, information dissemination, management science and engineering, emotion recognition.
Publications: more than 12 papers.

Suyu Dong, born in October, 1989 ,Liang Cheng, Inner Mongolia, China

Current position, grades: student at the College of Management, Inner Mongolia University of Technology.
University studies: Bachelor degree in Software Engineering from East China Jiaotong University in June 2012.
Scientific interests: Management science and engineering sentiment analysis.

Jing Yang, born in May, 1981, Feng Zhen, Inner Mongolia, China

Current position, grades: Associate Professor at the College of Management, Inner Mongolia University of Technology.
University studies: Doctor's degree in Economics from Beijing Central University of Finance and Economics in June 2012.
Scientific interests: internet public opinion, information dissemination, emotion recognition.
Publications: more than 8 papers.

Optimal model of highway road based on GPS

Huaping Li*

Chongqing City Management College, China Chongqing, 401331

Abstract

Intelligent transportation system (ITS) has gradually become reality with the increasingly wide application of GPS in car navigation system. It is wider and wider applied into traffic management to support short-tern traffic prediction, provide the information for public travel rapidly and effectively and realizes the effective supervision on traffic state. It can also relieve the pressure of highway road in our country. This paper focused on the study of floating car data technology based on GPS in ITS and developed relative procedure based on algorithm theory. It designed out a solution suitable for highway real time road condition system based on GPS floating car data technology after the demand analysis, functional decomposition and concept design on highway real time road condition system and best road and shortest time selection on driving by mathematical modelling so that achieve highway road optimization.

Keywords: GPS floating car data, GPS, ITS, mathematical modelling, driving time

1 Introduction

Developments of modern society bring people highly developed economy and wonderful material lives as well as negative effect such as traffic jam. City traffic infrastructure cannot keep up with traffic demand growth. City planning, land utilization and highway network planning are not reasonable. In addition, urban bus transport lags behind. And growth speed of mini car is too quick and volume of parking lot is not enough [1]. These problems all deteriorate traffic jam, which affects rapid development of social economy to a large extent. Therefore, scholars and experts from different countries begin to devote to this problem and some positive achievements have gained. Precondition for these achievement need improvement and development of advanced technology such as ITS and GPS [2].
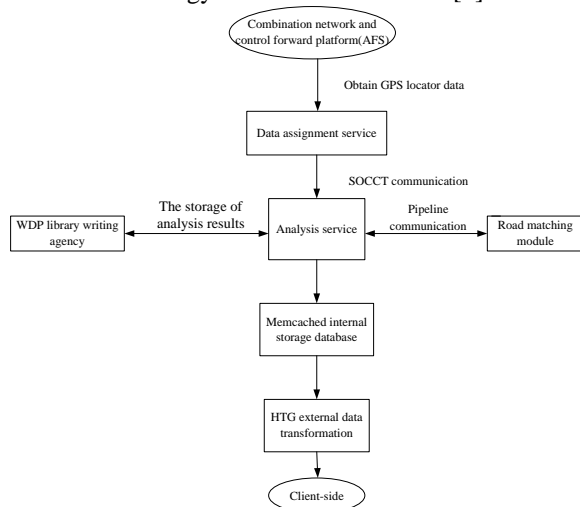
FIGURE 1 System overall flow figure

Although the research on ITS in our country started late, the research result on this aspect of experts is delightful. Zhou Yang from Nanchang Aviation University developed a national highway traffic system-B/S system based on GPS floating car data technology in research of highway real time road condition system based on GPS floating car data [3]. This system realizes 5 min real time refreshing and road condition prediction function. Meanwhile, it provides second development interface for external which is convenient for second development on other map developers. Bai Fengwei [4] made information more efficiency and diversified by improving highway information collection system in Opinion and Suggestion on Information Collection of Highway Monitoring System. Wang Junming [5] from Chang'an University studied topic of digitization of highway simulation monitoring system and laid a strong foundation for networking of highway monitoring system in research of highway monitoring system and relative device. This paper focused on study on principle of floating car data based on GPS and algorithm realization. And a highway real time road condition system was developed. Compared to traditional road condition system, it can collect road condition information flexibly with little cost. It is a good solution for current traffic jam.

2 Current situation overview and research thought of highway

Most road condition system in our country adopts fixed data collection means in view of current real time road condition system. For instance, Shanghai public transport system and Tianjin real time road condition all adopt video collection technology. And these systems are mainly applied in main traffic artery in city centre; however, road

---

Corresponding author e-mail: lhpcqcs@163.com

condition information collection by floating car data technology is still in theoretical research and verification stage in most area. Some cities have gradually applied floating car data technology to obtain real time road condition information.

Author programmed this paper according to development procedure of software system combining with basic requirement of master's thesis. The detailed flow figure is as Figure 1.

3 GPS floating car data technology and algorithm

3.1 GPS FLOATING CAR DATA TECHNOLOGY BASIS

3.1.1 Overall structure principle explanation

Floating car data technology based on GPS is based on GPS combining with geographical information technology, wireless communication technique such as CDMA, GSM, WCDMA, etc. computer communication and data processing technology [5]. It collects information through mobile car terminal installed in driving car and constructs computer model to process and analyse locator data by matching software in information centre. Thus, traffic jam condition information is obtained. The overall structure figure is as Figure 2.
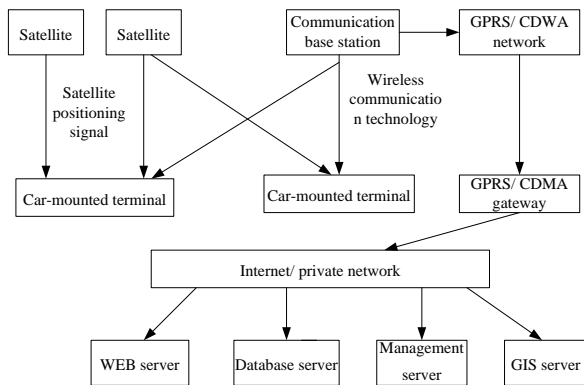
FIGURE 2 Overall structure chart based on GPS floating car data technology

3.1.2 Car-mounted terminal description

GPS car-mounted terminal is also termed as GPD car-mounted monitoring terminal. It monitors car position information and car state and provides software and hardware comprehensive system relying on satellite positioning, geographic information and wireless communication technique means. GPS car-mounted terminal is mainly composed of GPS positioning module, microprocessor unit, wireless communication module, client-side software and other extension unit [6]. The structure is shown in Figure 2.

3.2 MAP MATCHING ALGORITHM DESCRIPTION

Highway road condition system makes analysis on national highway road condition. Speed of matching algorithm largely affects the efficiency of system. Therefore, the critical part of matching algorithm is to match GPS information of floating car data to relative road as soon as possible. The basic flow is shown in Figure 3.
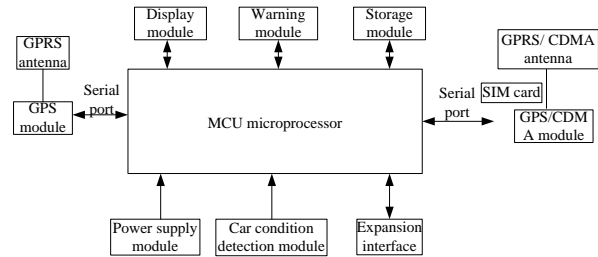
FIGURE 3 Matching algorithm flow chart

3.3 SEARCH AND REALIZATION OF MATCHING DISTANCE

In this system, the realization of matching algorithm is through the analysis from MAPX. And matching of GPS and highway road segment is realized through primitive lookup algorithm of MAPX. The function method is as follows:

OBJECT. Search within Distance (Source, double Distance, short Units, and short Search Type).

This method search surface features within a certain scope with some point or some primitive as standard. First parameter sets a point object or a feature object. Second parameter sets distance. Third parameter set unit. Fourth parameter sets return result and the position relationship between standard and distance [10].

Above function searches highway road segment information with certain scope by importing GPS position information. Then it analyses deeply angular separation according to initial matching road segment information and at last confirms the matching road segment of GPS point information.

3.4 ALGORITHM OF REAL TIME SPEED OBTAINING

Vehicle tracking method estimates average speed: one field in vehicle locator data packet. This speed is obtained by GPS car-mounted terminal. Its speed is also the instant speed of car. For the floating car data that has no continuous locating point in the same road segment, the instant speed of cars driving in that road segment is obtained and weighting for obtaining average speed.

$$w_r = \frac{p_r}{\sum_{r=1}^{n} p_r}, \qquad (1)$$

$$u_l = \sum_{r=1}^{n} w_r u_r, \qquad\qquad (2)$$

where w_r is the weight of floating car r, n is the number of floating car data positioning packet that pass road segment 1 and u_r is car speed in positioning packet.

The system is limited in locator data in single road segment. And national data size is too large. Therefore, this system adopts vehicle tracking - direct obtaining method to obtain speed of road segment in view of complexity of calculation.

4 Shortest driving time selection by mathematical modelling

4.1 MODEL BACKGROUND AND ANALYSIS

Road traffic state is nonlinear. And car, speed and road condition have interaction function. Its research is including micro model and macro model [8]. There are four indexed to measure in view of traveling of traveller. First is that expected driving time is shortest; second is the high reliability of route, that is, the possibility of serious block is low; third is that the journey is shortest; four is the lowest cost [9]. Sometimes these indexes are contradictory and balance is needed between two sides. In this paper, we took shortest driving time as design purpose.

4.2 PROBLEM ANALYSIS AND MODELING

We can conclude road condition on road L_i into three conditions: normal, jam and serious jam. The number of relative road segment is n_{i1}, n_{i2}, n_{i3}. Expectation and standard deviation of driving time on every road can be calculated according to the data obtained by sensor. Then the expectation and standard derivation of the whole road is:

$$\mu_{Li} = \sum_{i=1}^{n_{i1}} \mu_i^n + \sum_{i=1}^{n_{i2}} \mu_i^c + \sum_{i=1}^{n_{i3}} \mu_i^h, \qquad (3)$$

$$\sigma_{Li} = \sqrt{\sum_{i=1}^{n_{i1}} \sigma_i^n + \sum_{i=1}^{n_{i2}} \sigma_i^c + \sum_{i=1}^{n_{i3}} \sigma_i^h}, \qquad (4)$$

where $\mu_i^n, \mu_i^c, \mu_i^h$ are respectively expressed as the expectation of driving time in the condition of normal, jam and serious jam and $\sigma_i^n, \sigma_i^c, \sigma_i^h$ are respectively expressed as the standard derivation of driving time in the condition of normal, jam and serious jam. And obedience parameter of driving time of road L_i is normal distribution of $(\mu_{Li}, \sigma^2_{Li})$.

If the material is not sufficient, then there are three estimations on time: the most optimistic estimation, the most possible estimation and the most conservation estimation, which are respectively corresponding to normal, jam and serious jam. The three estimations can be recorded as a, b and c. And suppose that the probability of value as a, b, c are respectively $\frac{1}{6}$, $\frac{4}{6}$, $\frac{1}{6}$. Then expectation of finishing the road can be estimated by $ET = \frac{a + 4c + b}{6}$.

It can be known from Chebyshev in equation $p\{|t_i - ET| \geq \varepsilon \sigma_t\} \leq \frac{1}{\varepsilon^2}$ that when $\varepsilon = 3$, the possibility of deviation between time of finishing this road t_i and expected time ET exceed three times of divide difference of time need in finishing this road will not larger than $\frac{1}{9}$.

Suppose the minimum value of t_i is $a = ET - 3\sigma^t$ and the maximum value of t_i is $a = ET + 3\sigma^t$, then $b - a = 6\sigma^t$. Therefore, divide difference of time that need in finishing L_i can be estimated by $\sigma_t = \frac{b-a}{6}$.

Next we consider a road. Suppose that the road can be divided into m segments, then estimation on the expected time μ that needed in finishing this road can be expressed as sum of expected time of m segments. Estimation on time variance σ^2 can be expressed as sum of time variance of m segments road. Progressive obedience normal distribution of time can be known from central-limit theorem of probability theory. Based on it, probability P of finishing this road within a certain period of time can be calculated:

$$P = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\tau-\mu)^2}{2\sigma^2}} d\tau = \varphi\left(\frac{t-\mu}{\sigma}\right). \qquad (5)$$

According to above discussion, we can transform original problem into uncertain PERT network problem. Optimal purpose is defined as the shortest time needed in finishing some road in the condition of probability P.

5 Conclusions

This paper introduced ITS and mainly studied floating car data technology based on GPS in order to solve the increasingly serious traffic jam. Although research based on GPS floating car data technology have many theoretical achievement, many experiments focus on the area of urban main road where vehicles are intensive. However, these papers mainly face to national highway and the vehicles density and supporting points do not reach the amount of previous research achievement. Through practical observation and analysis, highway is different with previous research result. Vehicles density of highway is not such large as that of urban traffic road. But the structure of highway is simpler and few cars can reflect traffic condition. In addition, we increase jam feasibility for reference and explore a solution suitable for this system by nearby road detection and mathematical modelling.

References

[1] Yuxiao Z, Xianzhe L 2006 Reason and solution of urban traffic jam in our country *Hebei Traffic Science & Technology* **3**(3) 68-70 *(in Chinese)*

[2] Hongchuan D 2009 Developement situation and solution analysis of intelligent traffic system in our country *Jilin Taffic Science & Technology* (1) 60-3 *(in Chinese)*

[3] Yang Z 2012 Research on highway real time road condition system based on GPS floating car data *Nanchang Aviation University*

[4] Fengwei B 2012 Opinion and suggestion on information collection of highway monitoring system *Science & Technology Communication* **9**(17) 35-6 *(in Chinese)*

[5] Junming W 2013 Research on highway monitoring system and relative device *Chang'an University (in Chinese)*

[6] Golob T F, Recker W W 2004 A method for relating type of crash to traffic flow characteristics on urban freeways *Transportation Research Part A Policy and Practice* **38**(1) 53-80

[7] Bishop R 2004 Floating car data projects worldwide a selective review *Proceedings of ITS America Annual Meeting* 192-7

[8] Ryan A J 2011 Military applications of complex systems *Philosophy of Complex Systems* **37**(1) 723-80

[9] Hao Z 2009 Enlighten of Beijing intelligent traffic system on Shanghai Expo *Traffic & Transportation* **2** 14-6

[10] Qiang Z, Yao S, Lei W, et al 2010 Low cost GPS/ DR fault-tolerant integrated navigation system design *China Journal of Inertial Technology* **18**(4) 455-61 *(in Chinese)*

Author

Huaping Li, born in October, 1974, Chongqing Province, China

Current position, grades: associate professor.
University studies: Mathematics education, Chongqing Normal University in 1997.
Scientific interest: higher mathematics, mathematical modelling.

Information propagation in social network

Jianping Qiu[1*], Lichao Chen[1], Guifang Chen[2]

[1]*Institute of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China*

[2]*Computer Science and Technology Department, Taiyuan Normal University, Taiyuan 030012, China,*

Abstract

Information propagation network analysis provides a new way to investigate online activities. From the perspective of information propagation analysis we can understand it in a constant evolving way, that is, the content of the information is modified by the netizens with a certain probability during the whole propagation process. By analysing the online behaviour of netizens, we constructed an information propagation network on social networks. In this paper, we found that the original information can keep its influence on the netizens only when most of them are forwarders. Meanwhile this paper reveals influence propagation is aggregated, for example, netizens tend to give a low rating after a low rating, as well as a high rating following a high rating. Our findings are helpful in better understanding information propagation.

Keywords: information propagation, social network, scale-free network, power-law distribution, influence propagation

1 Introduction

With the rapid development of social networks, more and more people are using social networks to obtain information [1]. We no longer need to physically enter a library to obtain the latest news or to read a scholarly journal. A simple search through any computer or mobile device is enough to put at our disposal not only what we search for but also a trove of related findings that increase our curiosity and expand our horizons. Add to that the ubiquity of e-mail, instant messaging and Microblogs, and we find ourselves in a world of instant connectivity and potentially productive connections with social networks across the globe.

A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors, and is main tunnel of information propagation [2]. P. S. Dodds and D. J. Watts studied the accumulation effect of information propagation and E. Agliari et al took into consideration the degeneration of information on a spatial system. In social networks, e.g., scale-free networks [3], the information propagation process is much more complicated than the ordinary scenarios. The information changes constantly in its propagation process. The behaviour comes from the cumulative modifications during the propagation process however, whether the information can spread through the whole network depends not only on the existence of the connections among nodes, but also on their strategies. We divide the strategies into two types: one is to forward information directly; the other is to modify information before spreading it out. In this paper, we investigate information propagation in social networks, the netizens forwarding the information are named as forwarders. The netizens modifying the information are named as modifiers.

2 Related work

In 1967, Stanley Milgram is credited with introducing the notion of a small-world network to the social science community. Milgram's famous "six degrees of separation" experiment suggested that the distance between two people selected at random from the entire population of the United States is approximately six intermediaries. In 1999, Barabasi created another line of investigation with the invention of scale-free networks (non-random networks with hubs). In a number of studies of the structure of the Internet and WWW, Barabasi et al. discovered an emergent property of the decentralized Internet that it had emerged without central planning into a structure consisting of a small number of extremely popular sites called hubs, which have more influence, and a large number of "unpopular" sites with few links. Instead of being random, the Internet topology was very non-random. In fact, the probability that a site has k links obeys a power law, which drops off quickly for large k. Furthermore, they speculated that this was the result of a microrule called preferential attachment that the probability a site will obtain a new link is directly proportional to the number of links it already has. Thus, the more links a site has, the more it gets the so-called "rich get richer" phenomenon.

Figure 1 shows the flow web of the 1000 most-visited sites [4]. Solid circles represent websites and edges show information propagation flow. The size of circles is proportional to the logarithmic value of their flow. The red circles is web 2.0 sites, the blue ones is web 1.0 and the white ones is search engine. As we can see from Figure 1,

---

[*]***Corresponding author*** e-mail: 920868329@qq.com

netizens prefer web 2.0 to web 1.0.In Figure 2, the blue circles is statistics of the 1000 most-visited sites flow, the red line is analogue value of Zipf law. The reasons why there appear the phenomenon of a significant cut-off are lack of lower-ranking websites' data and rich-get-richer paradigm [5].As shown in Figure 2, the Internet is the power-law distribution which $\gamma < 1$.
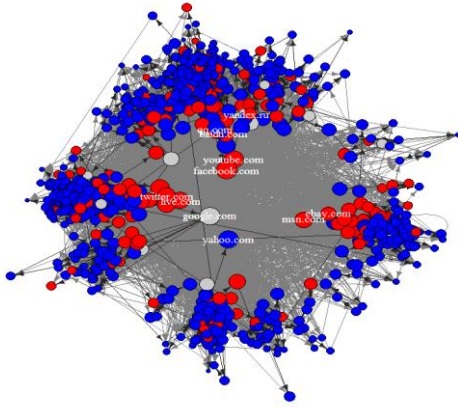
FIGURE 1 The network of top 1000 sites worldwide
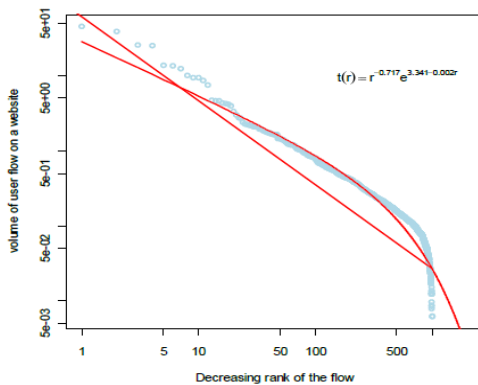
FIGURE 2 The flow distribution of top 1000 sites worldwide

3 Information propagation network

A information propagation network is a (directed or undirected) network, $G(t) = \{ N(t), L(t), f \}$, where $N(t)$ is a set of actors (nodes) with time-varying state, $L(t)$ is a set of influence degree ϕ for information, and f is a static mapping of $N \times N$ that defines the topology of G. Nodes and links have a value property that defines influence degree (for links) and position (for nodes) associated with a proposition. Further, let $S(t)$ be the state vector representing an actor's position, where $-1 \le S(t) \le 1$ and $S(t) = [s_1, s_2, \cdots, s_n]^T = $ *state vector of* G. The rate of change in the state of a netizen is dictated by the difference between the states of adjacent netizens, State Equation is as follows:

$$s_i(t+1) = s_i(t) + \sum_j [s_j(t) - s_i(t)]\phi^T = s_i(t) + \sum_j \Delta s_i(t) . \quad (1)$$

In this model, a netizen changes his or her position by an amount $\Delta S_i(t)$ after each interaction. A necessary (but not sufficient) condition for reaching a consensus in a communication network is $\sigma(L) < 0$, that is, a negative spectral gap, where σ is the largest nontrivial eigenvalue of L. At this moment, $\Delta s_i(t) = Ls(t)$, where L is the Laplacian of ϕ^T, as shown in Equation 2.

$$s_i(t+1) = s_i(t) + Ls_i(t) = [I + L]s_i(t) = [I + L]^{'} s_i(0) , \quad (2)$$

Our network can model the consequences of information propagation. Let s be a measure of stubbornness over netizen. In Figure 3, $s = 0$ means the netizen who disseminates the information directly to their neighbours denoted by a circle; $s \ne 0$ means the netizen delivers the information after modifying denoted by a square. So the content of the information is modified by the netizens with a certain probability during the whole propagation process.

When receiving a certain version of the information, the netizen (the receiver) becomes a disseminator (denoted by white colour in Figure 3) of the version. When receiving two or more different versions of the information, the netizen accepts the latest version. If a netizen receives the original information or the revised versions, which he sent before, they will turn to be a terminator (denoted by black colour in Figure 3). This is simply because once the disseminator has disseminated similar information to their neighbours, this means there is no need to do that again. On the other hand, the neighbours who send the versions of the information to them would not be interested in the similar information as well.
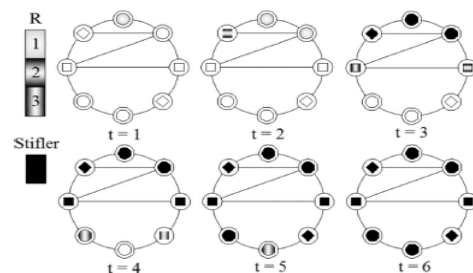
FIGURE 3 Information propagation network. *R* denotes the revised frequency of the information

4 Information propagation model

In our spreading model, netizens can play three roles: receivers, disseminators, and terminators, whose densities are denoted by $r(t)$, $d_x(t)$ and $e(t)$. Here, x denotes the version of information. The original version is one. We set the normalization condition $r(t) + d_{sum}(t) + e(t) = 1$ and $d_{sum} = \sum_x d_x$ .The role of a netizen starts with a receiver.

When information is injected into networks such as scale-free networks, the set of coupled properties can be

written as the following. Consider a receiver forwarder i with degree K after t steps. When receiving the information x, a probability with which it becomes a spreader d_x is shown in Equation (3):

$$P^{rd}_{r \to d_x}(t,k) = FkP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k \rangle}, \quad (3)$$

where $P(k)$ denotes the degree distribution of the networks and F denotes the forwarders' fraction. If i is a receiver modifier, the probability with which it becomes a modified information spreader d_{x+1} is shown in Equation (4):

$$P^{rd}_{r \to d_{x+1}}(t,k) = (1-F)kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k \rangle}. \quad (4)$$

The probability with which a disseminator d_x becomes a terminator e is shown in Equation (5).

$$P^{de}_{d_x \to e}(t,k) = kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t) + e_{k'}(t)\right]}{\langle k \rangle}. \quad (5)$$

We define $\langle R_k \rangle = \dfrac{\displaystyle\sum_{r \in \{r \mid \deg ree(r)=k\}} r_R}{N_k}$, where r_R denotes the last version of the information at netizen i before netizen i turns out to be a terminator. $\langle R_k \rangle$ represents the frequency that information has been modified on average at a netizen with degree K. The rate equation for the average revised frequency $\langle R_k \rangle$ on degree K can be written as Equation (6):

$$\frac{d\langle R_k(t) \rangle}{dt} = (1-F)P(k)kr_k(t)\sum_{k'}\frac{k'P(k')d_{k'}\langle R_{k'}(t) \rangle}{\langle k \rangle}. \quad (6)$$

The evolution of the densities $d_k(t)$ and $e_k(t)$ satisfy the following set of coupled differential formula:

$$\frac{dr_k(t)}{dt} = -kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k \rangle}, \quad (7)$$

$$\frac{dd_k(t)}{dt} = kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k \rangle} -$$
$$kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t) + e_{k'}(t)\right]}{\langle k \rangle}, \quad (8)$$

$$\frac{dd_k(t)}{dt} = kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k \rangle} -$$
$$kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t) + e_{k'}(t)\right]}{\langle k \rangle}. \quad (9)$$

To clarify the result of evolution, we run extensive simulations on scale-free networks for five different values of $F = 0.1, 0.3, 0.5, 0.7, 0.9$. We generate BA network which are generated by $m_0 = m = 3$. With the initial conditions $r(0) = \dfrac{N-1}{N}$, $d(0) = \dfrac{1}{N}$ and $e(0) = 0$, $k = 6$. We measure the distributions of R in Figure 4. We define $\Phi(R)$ as the number of netizens who were the disseminators of the information revised R times before the information vanishes. One can observe that the majority of the netizens are infected by the versions revised in Figure 4. This observation indicates that the original information can keep its influence on the netizens only when most of them are forwarders in the information propagation networks.
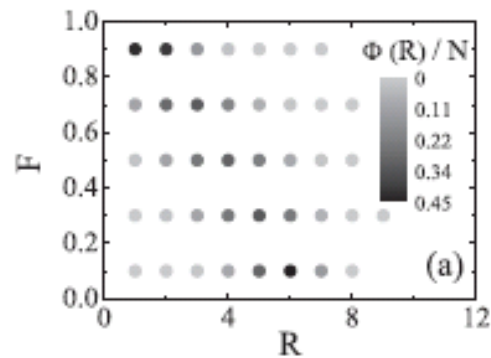
FIGURE 4 Simulation result

5 Influence propagation

There is tremendous interest in information propagation in social networks, fuelled by applications such as viral marketing, epidemiology, analysis of the spread of innovations, among many others. At the core of these applications there is a phenomenon called influence propagation, where actions performed by people propagate through a social network.

We collected data from MovieLens [6]. MovieLens is a movie rating system with five stars (i.e., ratings can be 1, 2, 3, 4 and 5). The dataset contains 6040 users, 3952 movies, and 1000292 ratings. From the dataset we construct a information propagation network, influence ϕ grows with the uploading index n. So we can obtain a simple formula $\phi = anY$, where a is a positive constant and Y is a multiplicative noise with mean one [7]. If a netizen's latest upload surpasses some "successful" threshold θ, she will continue to upload, otherwise she will stop [8]. Under these assumptions, a new upload with index n will fail to be successful with probability $P(anY < \theta) = P\left(\dfrac{aY}{\theta} < \dfrac{1}{n}\right)$. Let F be the Cumulative Distribution Function of the random variable $\dfrac{aY}{\theta}$. We see that a netizen who made n past uploads will stop at n with

probability $F\left(\frac{1}{n}\right) = F(0) + \frac{F'(0)}{n} + O\left(\frac{1}{n^2}\right)$, where $F'(0)$ is a positive constant. If the number of uploads follows a distribution P_n, which means a fraction P_n of netizens stop at n uploads. We can be written as $\frac{P_n}{\sum_{i=0}^{\infty} P_{n+i}} = \frac{F'(0)}{n}$.

Defining $G(n) = \sum_{i=0}^{\infty} P_{n+i}$, we get $-\frac{G'(n)}{G(n)} = \frac{F'(0)}{n}$. The solution to this equation is $G(n) \sim n^{-F'(0)}$, which implies that $P_n = -G'(n) \sim n^{-F'(0)-1}$. Therefore the distribution of the information propagation network is a power-law distribution.

To show the existing of influence propagation, we look at a case: we will vote with influence propagation after voting on some very high-quality or very low-quality object. In the absence of influence propagation, the next votes should be more or less the same as usual votes; while if the influence propagation exists, a vote will become the anchor of the next vote, and thus in average we will give high rating after voting on a high-quality object and low rating after a low-quality object.

We use the average rating to estimate a movie's quality, and to reduce the possible errors caused by personalized tastes and unreasonable votes, we only consider the objects getting more than ten votes. Although ratings cannot perfectly reflect qualities, they are correlated with qualities and can be naturally treated as anchors by netizens. For MovieLens, a movie is distinguished as low-quality or high-quality object if its average rating is lower than 2.0 or higher than 4.5.

Figure 5a shows the rating series of a netizen in MovieLens. We divide ratings into two kinds: one is positive and the other is negative, and show them without explicit values in Figure 5b, where we could find that

ratings in the same kind are aggregated. This kind of aggregation shows the influence propagation in voting behaviour, namely people are likely to give a high rating after a prior high rating while they are likely to give a low rating after a prior low rating. It is similar to the information propagation process that we are affected by neighbours in social network.
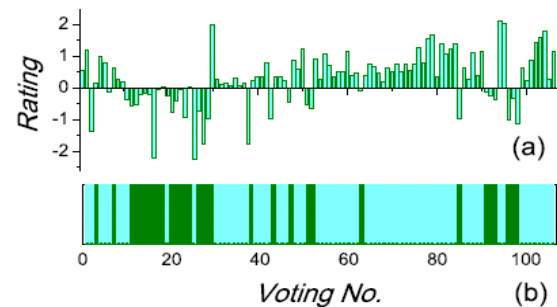
FIGURE 5 Rating series

6 Conclusions

Nowadays, Online social networking sites have become a popular way to share and disseminate content[9]. Their massive popularity has led to tremendous interest in information propagation. This paper presented information propagation network analysis. We showed the mechanisms of information propagation. As a result, we found that when the information spreads on the social networks the majority of netizens are influenced by the multirevised version. Meanwhile influence propagation is aggregated. Our result may provide a better understanding of information propagation.

Acknowledgments

References

[1] Dyagilev K, Mannor S, Yom-Tom E 2013 *Social Network Analysis and Mining* **3**(3) 521-41
[2] Manuel G R, Jure L, Bernhard S 2013 Modelling information propagation with survival theory *Proceedings of the 30th International Conference on Machine Learning* Atlanta USA June 2013 JMLR:W&CP 9-16
[3] Lou S L Yang X H 2013 Journal of Information Technology and Computer Science **6**(5) 10-5
[4] Google 2013 The 1000 most-visited sites on the web http://www.google.com/adplanner/static/top1000 2013

[5] David E, Jon K 2013 Networks, Crowds and Markets Reasoning about a Highly Connected World *Cambridge University Press*
[6] Grouplens 2013 MovieLens database http://www.grouplens.org 2013
[7] Roberts J, Hann I H, Slaughter S 2006 Journal of Management Science **52**(7) 984-99
[8] Oreg S, Nov O 2008 *Journal of Computers in Human Behaviour* **24**(5) 2055-73
[9] Zhang C J, Zeng A 2012 Phys A **391** 1822-10

Authors

Jianping Qiu, born in June, 1973, Taiyuan, China

Current position, grades: teacher of Computer in Institute of Computer Science and technology, Taiyuan University of Science and Technology
University studies: Institute of Computer Science and technology in Xian University of Science and Technology
Scientific interests: computer and social network.
Publications: 10 papers
Experience: teacher at Taiyuan University of Science and Technology since 2005.

Lichao Chen, born in March, 1961, Taiyuan, China

Current position, grades: Doctor of computer, professor in Taiyuan University of Science and Technology.
University studies: Institute of Computer Science and technology, Beijing University of Science and Technology
Scientific interest: computer science.
Publications: 50 papers.

Guifang Chen, born in January, 1957, Taiyuan, China

Current position, grades: professor in Computer Science and Technology Department, Taiyuan Normal University.
University studies: computer science.
Scientific interest: computer science.
Publications: 30 papers.

A complete solution for duplication detection over uncertain data

Peng Pan*, Xiaojun Cai

School of Computer Science and Technology, Shandong University, Jinan, P R China

Abstract

As the problem of uncertainty for duplication is increasingly prominent with the sharp growth of amount and scale for data sources, we need to pay more attention on it. However, the research on uncertainty about duplicated data is still on its start. In this paper, we propose a complete method for duplication detection with probability, which is efficient and suitable for large-scale dataset. Considering the large-scale background, firstly, we adopt the rapid cluster algorithm based on canopies to get blocks. Secondly, in order to generate the record sets, which represent entity, we provide one fuzzy cluster method over each block by assigning two thresholds. By doing these, we balance the complexity and accuracy. Finally, we assign the probability for each record in one block. The experiments show advantages over other present algorithms for performances.

Keywords: duplication detection, data uncertainty, canopy, data probability

1 Introduction

Duplication detection and data fusion are challenges in data integration. Duplication mainly arises from these situations: one entity has various expressions in different data sources owing to the heterogeneous schemas and semantics; two records might describe the different aspects for the same entity in one integrated data source, which has solved the problem of isomerism for schema and semantics. These issues will result in a large amount of fuzzy subordinations, which imply the uncertainty.

It is a tough task to handle data accurately for current methods of duplication detection and data fusion because one complete domain knowledge cannot be acquired easily, and the contents of duplicated records are usually inconsistent, incomplete and inaccurate. Therefore, manual interventions are necessary to improve the accuracy. However, it is not practicable for artificial means in large-scale data environments such as deep web. While automatic method is adopted to improve the efficiency, it usually choose the most possible information with the loss of some useful parts. These methods are not capable of guaranteeing the quality of duplication detection and data fusion.

As the problem of uncertainty for duplication is increasingly prominent with the sharp growth of amount and scale for data sources, we need to pay more attention on it. However, the research on uncertainty about duplicated data is still on its start. [1] builds a model for probabilistic database of duplicated data, and provides one query method based on the model. [2] provides a method for generating the probabilistic database over dataset. [3] proposes a algorithm for probabilistic duplication detection based on graph theory, but the complexity is so high that it is not suitable to be used in large scale dataset.

In this paper, we propose a method for duplication detection with probability, which is efficient and fit in large-scale dataset. Firstly, we design one algorithm for rapid blocking based on canopies to get a lot of block called canopy. Secondly, in order to generate the records set which represents an entity, we provide one fuzzy cluster method over each block by assigning two thresholds. Finally, we assign the probability for each record in one block.

The contributions for this paper are:
- Considering the large-scale background, we carry out the rapid cluster algorithm based on canopies, and then adopt fuzzy cluster method with two thresholds. By doing this, we balance the complexity and accuracy.
- We also provide one method to assign the probability for each record, whose experiment shows high efficiency.

2 Related works

For the uncertainty of duplication, it means that which records from different data sources are put together is uncertain, and what is the representative of one entity in one record set representing the entity is uncertain. For these uncertainty, [4] defines one "integration" operation to handle conflicting records, for example, the ages for one person in two relation are 23 and 24, respectively. The output is [23, 24], and each value has probability. [5] defines a data model to fuse the data tree expressed in XML, and assigns probability for each representative with one method called "frenquenlistic". The similar method

Corresponding author e-mail: ppan@sdu.edu.cn

appears in [6] and [7], where XML is assigned with probability, but the amount of representatives is reduced by outside domain knowledge. [8] provides a language to express the integrated results with uncertainty. [9] proposes a methods for generating probabilistic database over duplicated data. It also provides the algorithm to obtain the representative in the records set representing one entity, and assign one probability for each record. However, it does not explain how to obtain the original records. [10] generates one probabilistic database for duplicated data by hierarchical clustering with different parameters , and provides one effective query method over it. [3] provides the probabilistic duplication detection method based graph theory, but it has so high complexity that cannot be used in large scale data. [2] provides several algorithm for cluster to generate duplicated records set with uncertainty, and compares the methods for probability assignment.

3 Constructing possible sets over duplicated data clusters

3.1 BLOCK FOR MASSIVE DATA

For large scale datasets, it is inefficient for applying traditional cluster methods to construct huge matrix. Especially in real-time environments such as deep web query, the problem on how to improve the efficiency for cluster algorithms has been urgent. Recently, many methods such as Sorted Neighbourhood [11], Bigram Indexing etc. are proposed to solve the problem of large-scale cluster. In this paper, we adopt the idea of canopies [12] to improve the efficiency for duplication detection.

The process of blocking cluster data by canopies has two steps: firstly, one rough and low cost methods is applied to divide the source data into some overlapped subsets called canopies, whose certain data is the centre in the range. Secondly, canopies are one clustering algorithm with higher cost and calculations that are more accurate.

The main idea in this paper for canopies firstly find out all the data around one centre to create one canopy with minimal cost, then find the domain for next centre to create another canopy, this process iterates till all the data are included in canopies. Since canopies are overlapped, one data might exist in more than one canopy. Therefore, in order to guarantee all the data exist in canopies, we use two threshold: τ_1, τ_2 and $\tau_1 \geq \tau_2$. Algorithm 1 is the detailed description.

Algorithm 1: The rapid block methods based on canopies
INPUT: data source D OUTPUT: canopies
1 $CenterSet \leftarrow D$
2 $i = 0$
3 while $CenterSet \neq \phi$ do
4      $d \leftarrow \arg\min_{d \in CenterSet}(approxDist(d,d'))$
5      $canopy_i(d) \leftarrow \{ d' \mid d' \in D \wedge approxDist(d,d') \leq \tau_1 \} \cup \{d\}$
6      $CenterSet \leftarrow CenterSe - \{d'\mid d' \in \wedge approxDist(d,d') \leq \tau_2 \} \cup \{d\}$
7      $i = i + 1$
8 enddo

CenterSet is a candidate data centre point set, whose initial value is the whole data set. When *CenterSet* is null, the algorithm will end; the second to seventh lines describe the process for generating one canopy, and the fourth line select one data point d as the centre from two data, which form the shorted distance, and *approxDist* is the algorithm for rapid calculating the distances. The fifth line puts those data whose distances to the centre d are less than τ_1 into the canopy whose centre is d; the sixth line remove original centre and the data whose distances to the centre d are less than τ_2 from the centre data point set. Hence, these removed points are regarded as the centre points set of the canopy in this iteration, and this can guarantee each data point exists in only one centre points set of canopy. When *CenterSet* is null, it implies that all the data have been put into the canopies.

3.2 THE CLUSTER PROCESS IN BLOCKS

In each block formed by canopies, we adopt more fine clustering algorithm to generate cluster divisions. As a result, each division stands for one entity, the data indicating the same entity will be in one same cluster division. This paper divides each cluster division into two parts by two thresholds: core and edge. Among them, the core part is constituted by data with high similarity value, which is above the ceiling threshold θ_1 and the edge part consists of that with lower similarity value, which is between the bottom threshold θ_2 and ceiling threshold θ_1. Each data appears in one core part for only once, but can appear in more than one edge parts. Algorithm 2 is the detailed description.

Algorithm 2: The Clustering in canopy
INPUT: canopy S,
The similarity pair G in S ,
Thresholds: τ_1, τ_2
OUTPUT: final cluster divisions set C_f
1    $M \leftarrow G$
2    $C_s \leftarrow \emptyset, C_f \leftarrow \emptyset$
3    $CC \leftarrow \emptyset, CM \leftarrow \emptyset$
4   i=0
5   while $\max_{sim(w,v) \in M}(sim(w,v)) \geq \tau_2$ do
6        $u \leftarrow \arg\max_{u \in \{w\mid sim(w,v) \in M\}}(sim(u,v))$
7        $CC_i \leftarrow \{w \mid sim(u,w) \geq \tau_1 \wedge w \notin C_s\} \cup \{u\}$
8        $CM_i \leftarrow \{w \mid sim(u,w) < \tau_1 \wedge sim(u,w) \geq \tau_2 \wedge w \in S\}$
9        $M \leftarrow M - \{sim(u,v) \mid v \in CC_i \wedge sim(u,v) \in M\}$
10       $C_s \leftarrow C_s \cup CC_i$
11       $CC \leftarrow CC \cup \{CC_i\}$
12       $CM \leftarrow CM \cup \{CM_i\}$
13       $C_f \leftarrow C_f \cup \{\{CC_i\},\{CM_i\}\}$
14   i = i + 1
15   enddo

M is the table for similarity, which records the similarity value for all the data in one canopy; C_s is the core data nodes at present; CC is the core cluster set, whose

format is $\{\{CC_1\},\{CC_2\},\ldots\{CC_k\}\}$. CM is the edge cluster set, whose format is $\{\{CM_1\},\{CM_2\},\ldots\{CM_k\}\}$, C_f is the final cluster divisions, whose format is $\{\{\{CC_1\}, \{CM_1\}\}, \{\{CC_2\}, \{CM_2\}\} \ldots \{\{CC_k\},\{CM_k\}\}\}$.

The fifth to fourteenth lines describe the cluster process. The sixth line selects the pair(u,v) with maximum similarity from M, and set u as base. The seventh line puts all the data nodes whose similarities value with u are more than τ_1 into one core cluster division CC_i. The eighth line puts all the data nodes whose similarity with u is more than τ_2 and less than τ_1 into one edge cluster division CM_i. Since v and u have been put into core cluster division, the ninth line removes the similarity about v and u. The tenth and thirteenth lines update the sets C_s, CC, CM, C_f.

3.3 THE CALCULATION FOR PROBABILITY OF ELEMENTS IN CLUSTER DIVISIONS

The probability for one element in cluster division stands for the chance for which the element exists in one clean instance potentially. The method for calculating the probability has three steps:
1. Acquires the representative element *rep*.
2. Computes the sum d of distance between *rep* and each element in division.

3. Represents probability with $\dfrac{d}{\sum d}$.

Algorithm 3 is the description in detail.

Algorithms 3: The calculation for probability for elements in cluster divisions

INPUT：a set of records R,
            Cluster C over R,
            a similarity function $sim()$
OUTPUT：a set of probability P

1   for each $C_i \in C$ do
2          $C^* \leftarrow \varnothing$
3   for each $r \in C_i$ do
4                $rep = \arg\max_{r \in C_i}(\sum_{s \in C_i} sim(r,s))$
5   for each $t \in C_i$ do
6                      $p(t)= \dfrac{sim(t,rep)}{\sum_{r \in C_i} sim(r,rep)}$

The fourth line accomplishes step 1) and the fourth line finishes step 2) and 3). We adopt Softtf-idf [13] method with q-grams as the similarity function

4 Experiments

We conduct two experiments to evaluate the performances of the methods proposed in this paper:
    1) We compare the performances between the algorithm with canopy and one without canopy over the same dataset
    2) We evaluate the performances for various algorithms for probability assignments.
    We have collected 9978 book records from 50 online

book shop, and 1879 records is regard as the final dataset for experiments after manual tagging.

4.1 COMPARE THE PERFORMANCES BETWEEN THE ALOGIRTHM WITH CANOPY AND ONE WITHOUT CANOPY OVER LARGE SCALE DATASET

4.1.1 Experiment Design

We handle the dataset by the cluster algorithm with canopies and without canopies, where the latter is the method proposed in 3.2. We set the threshold   for the two algorithms to 0.75, and compare the precision ration, recall ratio and executing time over the results.

4.1.2 Evaluation Criteria

We given the calculating methods for precision ration and recall ratio.
    Suppose we have a exact cluster $G=\{g_1,\ldots g_k\}$ over relation R, let $C=\{c_1,\ldots c_k'\}$ is the k-th output cluster by clustering algorithm. We define a mapping function f from G to C, which maps each exact cluster g_i into one output cluster c_j, i.e. $c_j = f(g_i)$. Therefore, the precision ration and recall ratio for one cluster g_i are defined as following:
The precision ratio of a single cluster g_i:

$$Prec_i= \frac{|f(g_i)\bigcap g_i|}{|f(g_i)|}.$$

The recall ratio of a single cluster g_i:

$$Recl_i = \frac{|f(g_i)\bigcap g_i|}{|g_i|}.$$

As far a clustering algorithm as be concerned, its precision ration and recall ratio can be defined as the weighted average, which is defined as following:
    The precision ratio of all clusters:

$$Prec= \frac{|G \cap C|}{|G|} = \sum_{i=1}^{k}\frac{|g_i|}{|R|}Prec_i$$

The recall ratio of all clusters:

$$Recl= \frac{|G \cap C|}{|C|} = \sum_{i=1}^{k}\frac{|g_i|}{|R|}Recl_i.$$

In addition, we define the harmonic methods F_1, which is formulated as $F_1= \dfrac{2 \times Prec \times Recl}{Prec + Recl}$.

4.1.3 Results

Table 1 shows that the method with canopies has lower precision ratio and recall ratio than the algorithm without canopies, but has much lower execution time than the

latter. Taken together, the method with canopies is more preferable for large scale dataset.

TABLE 1 The performances compare

	Precision	Recall	F1	Execution Times(ms)
Canopies	0.709	0.964	0.817	3421
without Canopies	0.735	0.987	0.843	124852

4.2 PREFORMANCES FOR VARIOUS ALGORITHMS FOR PROBABILITY ASSIGNMENTS

4.2.1 Experiment Design

We compare the similarity measurements for Wjaccard, SoftTfIdf, Cosine w/tfidf by adopting the probability assignment proposed in 3.3. We randomly select ten clusters, and conduct these measurements respectively.

4.2.2 Experiment Criteria

We use the order parameter promise ration (*OPR*) to evaluate the influence of probability assignment over probability order. The calculation is following:

Suppose the right order for probability value in one records set is $L_{correct}$, we denote the probability value of record *r* as *p(r)*; the order for probability according to certain function is L_{output}, where the probability value of record r as *po(r)*. By computing the amount of pair (r_i, r_j) for which r_i and r_j appear in $L_{correct}$ and L_{output} together, we use the order parameter promise ratio (*OPR*) to evaluate the extent to which the probability assignment algorithm by one function retains the original order. The computing equation is $OPR = \dfrac{|(ri, rj) \mid r_i, r_j \in L_{output}, i \le j, p(r_i) \le p(r_j)|}{C_k^2}$,

where C_k^2 is all the pair of L_{output}.

4.2.3 Results

Figures 1 and 2 show that the *OPR* of SoftTfIdf method is close to Wjaccard, and higher than Cosine w/tfidf, and the execution time is far lower than the other two methods.

References

[1] Beskales G, Soliman M A, Ilyas I F, Ben-David S 2009 Modeling and Querying Possible Repairs in Duplicate Detection *Proceedings of the VLDB Endowment* **2**(1) 598-609
[2] Hassanzadeh O, M R J 2009 Creating probabilistic databases from duplicated data *The VLDB Journal* **18**(5) 1141-66
[3] Panse F, van Keulen M, Ritter N 2010 Indeterministic Handing of Uncertain Decisions in Duplicate Detection *Technical report* University of Twente (Netherlands) TR-CTIT-10-21
[4] Tseng F S C, Chen A L P, Yang W 1993 Answering heterogeneous database queries with degrees of uncertainty *Distributed and Parallel Databases* **1**(3) 281-302
[5] van Keulen M, de Keijzer A, Alink W 2005 A probabilistic XML approach to data integration *Proceedings of the 21st International Conference on Data* Engineering 2005 ICDE 2005 459-70
[6] Hunter A, Liu W 2006 Fusion rules for merging uncertain information *Information Fusion* **7**(1) 97-134

FIGURE 1 OPR for various algorithms for probability assignments
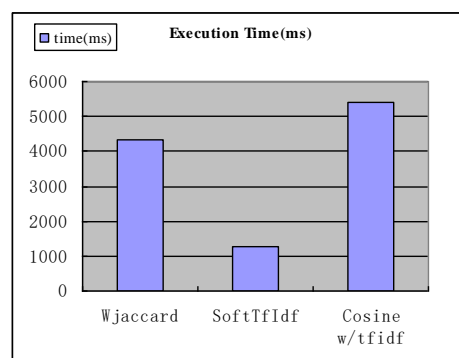
FIGURE 2 Execution Time for various algorithms for probability assignments

5 Conclusion

In this paper, we propose a complete method for duplication detection with probability considering the large scale background, firstly, we carry out the rapid cluster algorithm based on canopies. Secondly, in order to generate the records set which represents an entity, we provide one fuzzy cluster method over each block by assigning two thresholds. By doing these, we balance the complexity and accuracy. Finally, we assign the probability for each record in one block. The experiments show advantage over other present algorithms for performances.

[7] Hunter A, Liu W 2006 Merging uncertain information with semantic heterogeneity in XML *Knowledge and Information Systems* **9**(2) 230-58
[8] Calì A, Lukasiewicz T 2006 An approach to probabilistic data integration for the semantic Web *Uncertainty Reasoning for the Semantic Web I Lecture Notes in Computer Science* **5327** Springer-Verlag Berlin Heidelberg 52-65
[9] Andritsos P, Fuxman A, Miller R J 2006 Clean Answers over Dirty Databases:A probabilistic Approach *Proceedings of the 22nd International Conference on Data Engineering* 2006 30
[10] Gupta R and Sarawagi S 2006 Creating probabilistic databases from information extraction models *Proceedings of the 32nd international conference on Very large data bases* 965-76
[11] Baxter R , Christen P, Churches T 2003 A comparison of fast blocking methods for record linkage *ACM SIGKDD Workshop on Data Cleaning, Record Linkage, and Object Identification* 25-7

[12] McCallum A, Nigam K and Ungar L H 2000 Efficient clustering of high-dimensional data sets with application to reference matching *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* 169-178

[13] Cohen W, Ravikumar P, Fienberg S 2003 A comparison of string distance metrics for name-matching tasks *Proceedings of International Joint Conference on Artificial Intelligence* 73–78

Authors

Peng Pan, born in January, 1974, Anqiu, Shandong, China

Current position, grades: Lecture, Doctor of Computer Science.
University studies: Ph.D degree from Shandong University in 2010.
Scientific interest: deep web, data uncertainty, electronic commerce.
Publications: about 20.

Xiaojun Cai, born in September, 1976, Yizheng, Jiangsu, China

Current position, grades: lecturer, Doctor Candidate.
University studies: Shandong University.
Scientific interest: embedded technology, electronic commerce, data management.
Publications: about 10.

Towards an organizational learning framework for IS development

Hsin-Ke Lu*, Peng-Chun Lin

Institute of Information Management, School of Continuing Education, Chinese Culture University, Taipei, Taiwan (R.O.C)

Abstract

Information systems (IS) development has been one of IT managers' concerns for decades. To improve the effectiveness of IS development, many studies argued that development methods need the continuous learning mechanism to better utilize the knowledge along with all development stages. In this study, we discoursed the essential important of organizational learning for development team to deal with the complexities of IS development. The development team failed to learn from the experience in prior IS development projects because of limits of individual learning. Knowledge management concepts are gradually mature and will be fundamental theories to integrate the organizational learning mechanism into IS development. This paper illustrated all related literature and theories and proposed an organizational learning framework for IS development. The framework could aid the development team to recognize the value of organizational learning in the IS development activities and guide for devising the mechanisms for acquiring, maintaining and transferring that knowledge.

Keywords: IS development, organizational learning framework, knowledge management

1 Introduction

The rapid development of Information Technology (IT) increased enterprises' difficulty in the planning, development and management of Information System [1]. SIM had conducted a series of studies on the IT managers concerns in order to analyse the evolvement of the significant issues; in fact, this also represented enterprises' needs derived from their adaptation to environmental changes [2-5]. To facilitate further analysis, Niederman et al., [6] classified the issues along three dimensions and categorized them into four groups:

1) Business Relationship Issues: Including issues of data resources, IT and business strategy alignment, strategic planning, investigation of the role and contribution of IT, organizational learning, competitive advantage, and business process redesign etc.

2) Internal Effectiveness Issues: Including issues of system development, IT human resources management, measurement of IT effectiveness and system transfer etc.

3) Technology Infrastructure Issues: Including issues of information infrastructure, distributed systems, enterprise architecture, and communication networks etc.

4) Technology Application Issues: Including issues of end-user computing system, computer-aided software engineering, executive/decision system, office automation, and group support etc.

The study summarized the managers concerns on IS planning and development over the years and found out that IT and business strategy alignment was one core issue that received most attention from IT managers. Within recent 8 years, it ranked the top for 6 times. Nevertheless,

issues related to it including business process redesign, strategic planning of information technology and enterprise architecture also received much attention [3-5]. IS development was reported as another key issue of internal effective group. Improving the effectiveness of IS Development, as a core concern of IT managers, had varied in importance since the studies in the early 1980 [2]. The emergent technologies kept developers on a steep learning curve. High volume of IS development backlog and dynamic nature of development also contributed to this core issue. Lyytinen and Robey [7] noted IS Development projects remain susceptible to failures because organizations fail to learn from their own experiences. Therefore, to improve the effectiveness of IS development, IS development methods need the continuous learning mechanism to better utilize the knowledge and integrate new technologies into the IS development.

2 Learning perspective on IS development

IS development methods were designed to be suitable for different environment and purposes. In general, IS development is a set of iterative procedures, which involved continuous change and recycling. This enables the flexibility of IS structures and functions toward the dynamic context. Beside the different operational logic of software engineering methods, the developers have to face the problems caused by the implementation in different environments. The skills of the problem solving mostly are based on their related experiences. Therefore, integrating the mechanism of organizational learning into the

---

* *Corresponding author* e-mail: sklu@sce.pccu.edu.tw

development process could not only support the problem-solving ability but also enhance the development effectiveness.

IS development in fact was a series of continuous development process (the development cycle and maintenance recycle) and not a single cycle of system development activities. This study emphasized that for the IS development and maintenance process, we needed organizational learning to share the experiences from different development stages. Hayes [8] noted that the learning organizations constantly strive to be better, placing great emphasis on experimentation, integration, application and the building of critical organizational capabilities in all operations. That the effectiveness of each development project will gradually improve through learning behaviours will be a challenge in the future and a critical factor to maintain its effectiveness. This perspective showed the potential value of learning mechanism.

It is essential for IS development to upgrade the individual learning to organizational learning. Organizational Learning is mediated by the learning of individual of organizational members. The difference between them seemed to be varied on different research positions. Popper and Lipshitz [9] clarified the distinction between the two conceptions of individual learning in organizations and learning by organizations, which was also called organizational learning. To maximize the benefits of organizations, it was necessary that individual learning was extended to organizational learning for organizations to systematically collect, organize, and share

the valuable experience and knowledge [10]. Lyytinen and Robey [7] also mentioned that IS development was a complicated project activity involved with many team members' collaboration. The nature of learning behaviour (individual and organizational levels) was to extract, save, integrate, and pass these related experiences and adjust or increase habitual domain by feedback information.

3 IS Development and learning mechanism

In the IS development context, the professionals also need to learn the problem solving in situations where technical and contextual information are inseparable. The traditional IS development methods put much emphasis on the management of development document, but IS development methods of learning orientation put much emphasis on the organization and sharing of accumulated experiences in addition to the management of development document. The main difference was that the latter one must be based on the results of the former one, and of course this must satisfy cost consideration; because if the results of the former one couldn't assist the effectiveness of the redevelopment procedure and it couldn't be assisted by the former experiences, and then it would be greatly restricted by the former system. However, system development methodology needed to interact with the sub-models of organizational learning and hence the activities of the experience of introduction, the evaluation and synthesis of experience and feedback of application information must be added in each stage.
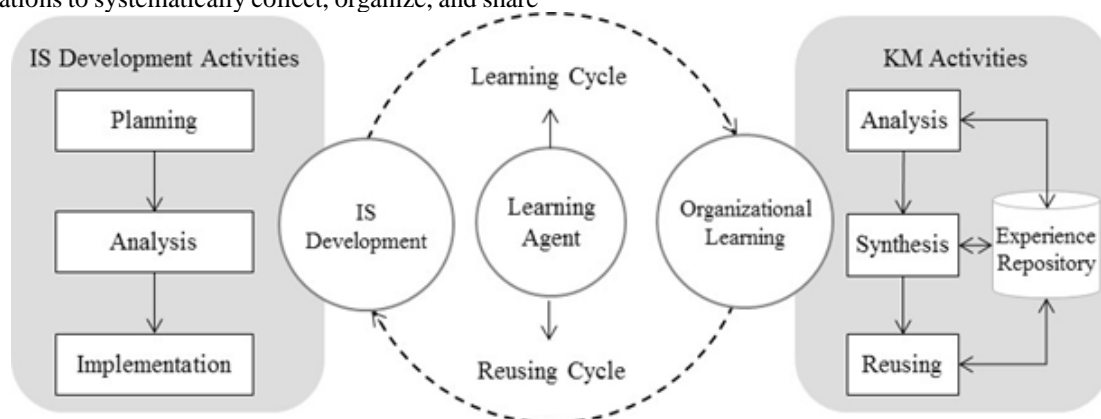
FIGURE 1 Organizational learning oriented IS development model

IS development methods by which developers accumulate knowledge about action-outcome relationships and the effect of the environment on these relationships [11]. Organizational learning oriented IS development model (Figure 1) was to formalize learning behaviours inside the system development process (individual learning level) and make them a part of organizational learning (elevated to organizational learning level). This would make a development method have the ability to learn and evolve; the main part of this model was a learning sub-model, and it enhanced the potential ability of a development sub-model and further improve the overall effectiveness of a development sub-model through the

analysis, refinement, integration and reuse of related experiences of a development sub-model. Furthermore, regarding IS development as learning system from which the team members could learn and overcome the inherent biases established among the professionals.

3.1 EXPERIENCES AS KNOWLEDGE ELEMENT

The existed development methods for IS majorly focus on the system documents and artifacts management for collaboration in the development context. It causes that IS development organizations tend not to view development activities as organizational learning from prior experiences;

rather, as documents management. In the IS development context, the knowledge to learn are a set of assumed causal relationships between actions taken during IS development stages and desired outcomes [7]. In addition to considering the development documents and artifacts, this study placed direct attention upon the lesson-learn knowledge, named experience-based knowledge elements. By doing so, the IS development is able to adjust its actions based upon knowledge elements gained from internal and external experiences.

For the experiences processed in learning model including action-outcome relationship in system development stages and the contextual information, these experiences were required to be communicable, consensual and integrated. This type of relationship originated from the development process and outcome of IS development methods. Saving these relationships would assist developers' learning about these methods. Martin [12] used encyclopaedia to archive the experiences of action-outcome relationship produced in the

development process to support the overall development cycle.

3.2 KNOWLEDGE MANAGEMENT MODEL FOR ORGANIZATIONAL LEARNING

Utilizing the mapping of knowledge management mechanism could present how organizational learning theory could be introduced into the IS development stages and also reflect what learning activities should be included in the IS development method with learning mechanism [13, 14]. Organizational learning mechanism could be simply divided into demand loop, supply loop, and feedback loop; using these 3 loops could assist the process of knowledge creation, accumulation and sharing. This study integrated Moore's learning cycle theory [15] and knowledge management model and used this theoretical structure as the basis to construct a new organizational perspective on IS development (Figure 2).
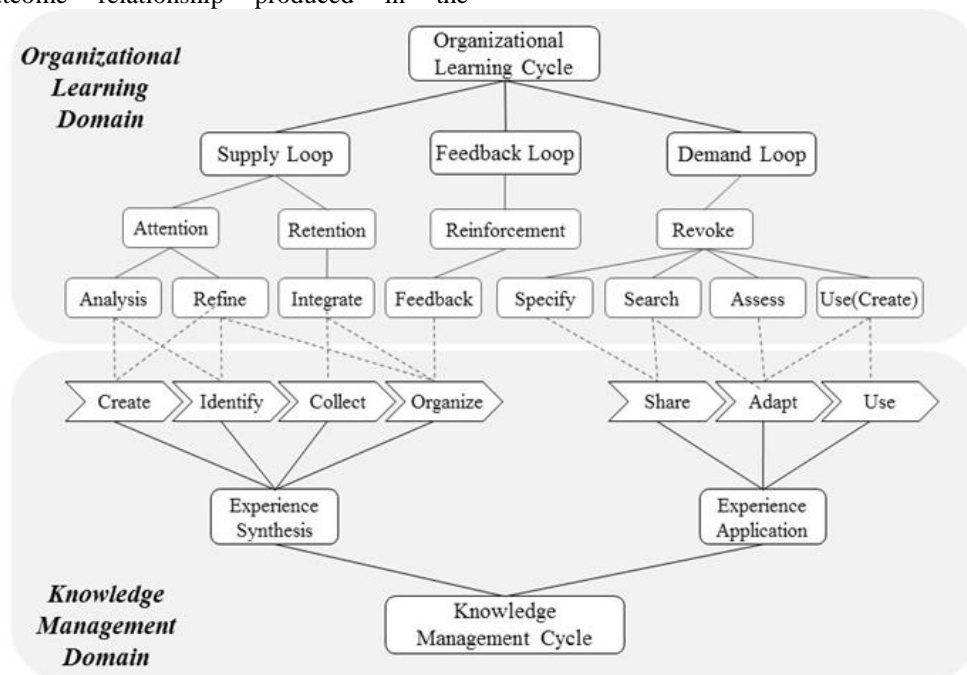
FIGURE 2 Mapping of knowledge management and organizational learning

1 Supply Loop:
   This loop could complete learning activities of knowledge attention and retention.
   - Analysis: analysing the value of knowledge;
   - Refine: refining knowledge and make it more generalized;
   - Integrate: integrating knowledge into knowledge repository.
2 Demand Loop:
   This loop could revoke knowledge for learning purpose.
   - Specify: specifying knowledge in demand;
   - Search: searching knowledge satisfying the demand;
   - Assess: assessing knowledge found;
   - Create: applying knowledge to future context.
3 Feedback Loop:

This loop could reinforce learning activities; its main purpose was to provide feedback about the status of knowledge introduction to be the basis of repository renewal so that its adjustment and activation could improve the core competence of organizational learning.

4 Organizational learning framework for IS development

In general, IS development is considered as a systematic and complex activity towards specific project objectives. It can be enhanced by an experience-repository for helping one developer learn from the others in difference stages of IS development. Shortening learning curve for IS development knowledge would be one of critical success

factors [7]. To implement learning mechanism into IS development cycle, knowledge management is a fundamental concept from which to build it up. Sanchez [16] noted that knowledge management is much more effective at leveraging the organizational knowledge and also become better at systematically driving organizational learning.

We proposed an organizational learning framework for IS development in order to improve IS development with learning mechanism. The learning mechanism was divided into experience reusing cycle and experience synthesis cycle. Using these two cycles enabled the development stage to contain learning activity; i.e. experience reusing cycle and experience synthesis cycle were the important tasks to make development activity to have learning mechanism. Utilizing the interactive model of IS development could present how organizational learning theory could be introduced into IS development and also reflect what learning activities should be included in the development stages with knowledge management mechanism (Figure 3).
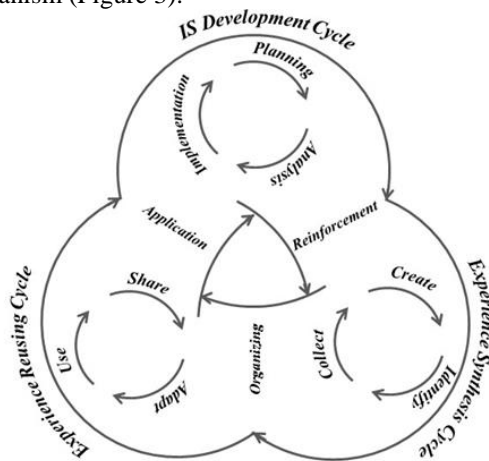
FIGURE 3 Organizational learning framework for IS development

The organizational learning framework of IS development addressed more on the experience synthesis and reusing to make the value of learning from experience.

1. Experience Synthesis Cycle:

Nonaka [17] noted that organizational knowledge was created as involving a contextual interplay between the tacit and explicit dimensions of knowledge and a synthesis flow as knowledge moves through individual and group levels. In the IS development context, experience is identified and synthesized from different IS development stages. The IS development group would perform

synthetic analysis and summary in the group meeting about the problem cases and related solutions accumulated by the developers (as learning agents). The feedbacks collected after using related experiences were saved in organizational repository to enhance or reduce the cause-and-effect degree of the experience in cause-and-effect relationship model. Fu and Anderson [18] proposed a reinforcement learning mechanism to provide an integrated explanation of enhancing the recognition level of knowledge.

2 Experience Reusing Cycle:

An important aspect of the knowledge management was reusing the knowledge rather than knowledge itself [19]. The reusing cycle was spread in the development stages. After a problem emerged in the development activity, developers would reuse the related experience.

Besides, utilizing formal technology of experience extraction and presentation model could integrate these experiences into the basic structure of IS development methods. By doing so, it would formally transfer individual learning activity to organizational learning activity. It also can further improve the performance of IS development methods.

5 Conclusions

For improving the IS development, some researchers claimed to advance development technologies. Some identified a critical reason for IS development failures, an absence of organizational learning applied to the complexities of IS development activities. The existed development methods for IS majorly focused on the system documents and artifacts management for collaboration in the development context. It caused that IS development organizations tend not to view development activities as organizational learning from prior experiences; rather, as documents management. Many different IS development methods had far distinct rules and logics under different context and personnel quality, and if we could integrate and share the development experiences of these methods so that they could promote the organic growth of organizational level and assist future tasks of system development and maintenance, and this would finally build up significant resources for organizations. Finally, an organizational learning framework for IS development was proposed. For IS development methods with learning mechanism, all related experiences in the process needed to be identified and archived for reusing in the recycling the IS development stages.

References

[1] Farwell D W, Kuramoto L, Lee D, Trauth, E M, Winslow C 1992 *Information Systems Management* **9**(2) 7-14
[2] Brancheau J C, Janz B D and Wetherbe J C 1995 in MISRC
[3] Luftman J N 2004 *MIS Quarterly Executive* **4**(2) 270-286
[4] Luftman J N and BenZvi T 2010 *MIS Quarterly Executive* **4**(9) 263-73
[5] Luftman JN and Derksen B 2012 *in CIOnet*

[6] Niederman F, Brancheau J C and Wetherbe J C 1991 *MIS Quarterly Executive* **15**(4) 475-500
[7] Lyytinen K, Robey D 1999 *Information Systems Journal* **9** 85–101
[8] Hayes R H, Wheelwright S C, Clark K B 1988 *Dynamic Manufacturing: Creating the Learning Organization*
[9] Popper M, Lipshitz R 1998 *Journal of Applied Behavioral Science* **34**(1) 61–78

[10] Nolan Norton Institute 1998 *Putting the Knowing Organization to Value*

[11] Rau D, Haerem T 2010 *Information Systems Frontiers - ISF* **12**(3) 287-97

[12] Martin J 1995 *Technology, Strategy Amacom*

[13] Lorino P 2001 *Knowledge Management and Organizational Competence* 177-209

[14] Argote L 2013 *Organizational Learning: Creating, Retaining and Transferring Knowledge*

[15] Moore J M, Bailin S C 1991 *Domain Analysis and Software Systems Modeling*

[16] Sanchez R 2005 *Knowledge Management and Organizational Learning: Fundamental Concepts for Theory and Practice*

[17] Nonaka I 1994 *Organization Science*, **5**(1) 14-37

[18] Fu W T, Anderson J R 2006 *Journal of Experimental Psychology: General* **135**(2) 184-206

[19] Alavi M, Leidner D E 2001 *MIS Quarterly Executive* **25**(1) 107-36

Authors

Hsin-Ke Lu, born on April 20, 1968, Taiwan(R.O.C.)

Current position: ass. professor LU, Hsin-Ke is the CEO of the School of Continuing Education, the Director of Information Management (Master Programs for Working Learners) at Chinese Culture University in Taiwan.
University studies: information management, National Taiwan University.
Scientific interest: e-learning, enterprise architecture, information system planning, business dynamics, and lifelong learning.
Publications: 45 papers.

Peng-Chun Lin, born on May 18, 1978, Taiwan(R.O.C.)

Current position: instructor of Information Management Department at Chinese Culture University in Taiwan.
University studies: Information and Computer Education, National Taiwan Normal University
Scientific interest: e-learning, social network, simulation learning, optimal network management and recently on enterprise architecture
Publications : 25 papers.

Robust face and facial feature localization using the dual skin model

Wei Li*

Computer School, China West Normal University, Shida Rd. 1, Nanchong, Sichuan, China

Abstract

A fast and adaptive face and facial feature localization algorithm for colour images with sophisticated background is present. In this algorithm, a self-adaptive pre-processing method was provided to depress the colour bias and the high light. Then the CbgCbr-YIQ dual skin model was proposed to acquire the integrated skin similarity for improving the quality of skin segmentation and extraction. After the morphological post-processing, by using the Adaboost classifier and the information of spatial position, the facial feature positioning was fast realized finally. Experimental results showed the robustness and good performance of the proposed algorithm.

Keywords: pre-processing, dual skin model, facial feature localization, face recognition, Adaboost

1 Introduction

In research, fields such as image understand and computer vision, as an important stage or a key problem in face recognition, the face and facial feature localization often has received significant attention. Variety of methods were proposed to solve the face detection problem, such as the method based on template matching [1], method based on classifier [2-5], method based on colour and space features [6,7]. For the first method, it may be spent too much time on the calculation of matching. For the second method, the accuracy and stability of detection is greatly affected by the trained classifier and chose samples. For the third method, the accuracy of facial feature localization depends on the built skin model and the results of skin segmentation. The algorithm in this paper combined the advantage of classifier and advantages of colour and space features. First, an adaptive pre-processing method was proposed to improve the quality of input image. Second, a new CbgCbr-YIQ dual skin model was built to acquire the integrated similarity for skin segmentation and extraction. Then the face candidates came from the hole-filled skin-extracted image was got by morphological post-processing. Finally the face candidates based on the dual skin model were regarded as the testing objects of Adaboost classifier, and the spatial information of facial features were jointly used to localize the facial features for the face image.

2 Algorithm flow

As shown in Figure 1, the algorithm is consisted of five stages:
1. Pre-processing: use pre-processing to depress the colour excursion and high light for improving the quality of original image (Figure 1a). Accordingly the pre-processed image can be acquired (Figure 1b).
2. Building the dual skin model and acquiring integrated similarity: the CbgCbr-YIQ dual skin model consisted of the CbgCbr ellipse model and the adaptive YIQ model. It aimed to combine features of two models. Figures 1c and 1d showed the corresponding skin-similarity images based on CbgCbr and YIQ models. In this stage, the integrated similarity image can be acquired (Figure 1e) based on the CbgCbr-YIQ dual skin model.
3. Skin segmentation and extraction: based on integrated similarity, skin-segmented and skin- extracted images can be acquired by binary segmentation and extraction, as shown in Figures 1f and 1g.
4. Post-processing: use morphological technology contained the noise reduction and hole filling to realize the post-processing and prepare for the subsequence localization. With the noise reduction technology in post-processing, the denoised skin-segmented image and denoised skin-extracted image can be acquired (Figures 1h and 1i). With the hole filling technology, the hole-filled skin-segmented image and hole-filled skin-extracted image can be acquired (Figures 1j and 1k). The denoised skin-segmented or skin-extracted image decreased the noise disturbance but preserve the facial features. The hole-fill skin-segmented image or skin-extracted image contained the face candidates' targets.
5. Positioning face and facial features: Use Adaboost classifier to detect face targets in face candidates' targets. Accordingly, the hole-filled face-segmented image and hole-filled face-extracted image can be acquired (Figures 1m and 1l). Besides the denoised face-segmented image and denoised face-extracted image can be acquired (Figures 1n and 1o). By difference and combined using the

---

Corresponding author e-mail: nos036@163.com

spatial information of facial features, the facial features can be localized (Figures 1p, 1q and 1r).
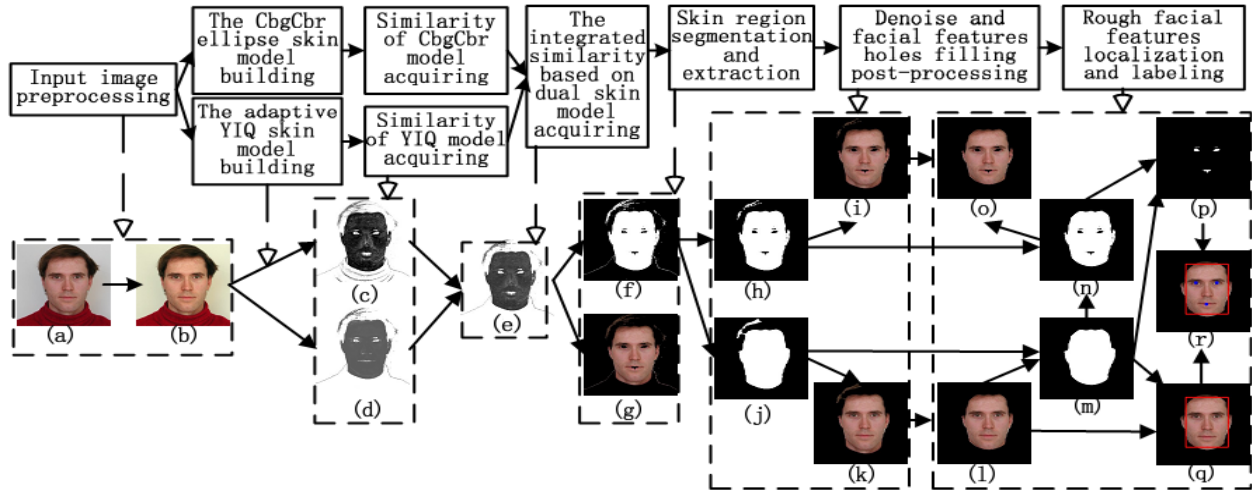
FIGURE 1 The algorithm flow: a) input image, b) pre-processed image, c) skin-probability image of CbgCbr ellipse model, d) skin-probability image of YIQ model, e) integrated probability image of dual skin model, f) skin-segmented image based on dual skin model, g) skin-extracted image based on dual skin model, h) denoised skin-segmented image, i) denoised skin-extracted image, j) hole-filled skin-segmented image, k) hole-filled skin-extracted image, l) hole-filled face-extracted image, m) hole-filled face-segmented image, n) denoised face-segmented image, o) denoised face-extracted image, p) feature image, q) face-localized image, r) face and facial features localized image

3 Pre-processing

The quality of colour input image is often affected greatly by input device and environment. The problems of colour excursion and high light are often caused by sophisticated background and variant light conditions. For improving the quality of colour input image and helping to the subsequent operations (e.g. skin segmentation and extraction), the proposed pre-processing in this paper aimed to adjust the colour and brightness adaptively to depress the colour excursion and the high light. The main steps of pre-processing are as follows:

Step 1: Accumulate separately $R_{i,j}$, $G_{i,j}$, $B_{i,j}$ and $Y_{i,j}$ components of every pixel to calculate their average values: aR, aG, aB and aY using Equation (1):

$$\begin{cases} Y_{i,j} = 0.299 \times R_{i,j} + 0.578 \times G_{i,j} + 0.114 \times B_{i,j}, \\ aR = \dfrac{\sum\limits_{j=0}^{H}\sum\limits_{i=0}^{W} R_{i,j}}{(H \times W)}, \ aG = \dfrac{\sum\limits_{j=0}^{H}\sum\limits_{i=0}^{W} G_{i,j}}{(H \times W)}, \\ aB = \dfrac{\sum\limits_{j=0}^{H}\sum\limits_{i=0}^{W} B_{i,j}}{(H \times W)}, \ aY = \dfrac{\sum\limits_{j=0}^{H}\sum\limits_{i=0}^{W} Y_{i,j}}{(H \times W)} \end{cases} \quad (1)$$

where, H and W are the height and the width of the input image, i and j are a point's abscissa and ordinate of the input image.

Step 2: Use aR, aG, aB and aY in Equation (2) to calculate the adjustment coefficients: cR, cG and cB.

$$cR = \frac{aY}{aR}, cG = \frac{aY}{aG}, cB = \frac{aY}{aB} . \quad (2)$$

Step 3: Using Equation (3), use $R_{i,j}$, $G_{i,j}$, $B_{i,j}$ and $Y_{i,j}$ components of every pixel to calculate light adjustment coefficients: $lR_{i,j}$, $lG_{i,j}$, $lB_{i,j}$.

$$\begin{cases} lR_{i,j} = \left[a\tan\left(\dfrac{R_{i,j}}{Y_{i,j}}\right) + a\tan\left(\dfrac{R_{i,j}+G_{i,j}+B_{i,j}}{R_{i,j}\times 3}\right) \right] \times 0.5 \\ lG_{i,j} = \left[a\tan\left(\dfrac{G_{i,j}}{Y_{i,j}}\right) + a\tan\left(\dfrac{R_{i,j}+G_{i,j}+B_{i,j}}{G_{i,j}\times 3}\right) \right] \times 0.5 \\ lB_{i,j} = \left[a\tan\left(\dfrac{B_{i,j}}{Y_{i,j}}\right) + a\tan\left(\dfrac{R_{i,j}+G_{i,j}+B_{i,j}}{B_{i,j}\times 3}\right) \right] \times 0.5 \end{cases} \quad (3)$$

Step 4: By Equations (4) and (5), use the colour adjustment coefficients and the light adjustment coefficients to adjust the $R_{i,j}$, $G_{i,j}$, $B_{i,j}$ components and acquire the adjusted $R'_{i,j}$, $G'_{i,j}$, $B'_{i,j}$ components:

$$\begin{cases} R'_{i,j} = R_{i,j} \times cR \times lR_{i,j} \\ G'_{i,j} = G_{i,j} \times cG \times lG_{i,j} \\ B'_{i,j} = B_{i,j} \times cB \times lB_{i,j} \end{cases} \quad (4)$$

$$\begin{cases} R'_{i,j} = 255 \ \ if\,(R'_{i,j} > 255) \\ G'_{i,j} = 255 \ \ if\,(G'_{i,j} > 255) \ . \\ B'_{i,j} = 255 \ \ if\,(B'_{i,j} > 255) \end{cases} \quad (5)$$

Step 5: Normalize separately the $R'_{i,j}$, $G'_{i,j}$, $B'_{i,j}$ components of input image to the range [0 255].

Figure 2 shows an example of pre-processing. Comparing Figures 2a and 2b, Figures 2g and 2h, Figures 2i and 2j, note that the colour bias and the high light were depressed partly. Comparing Figures 2c and 2d, Figures 2e and 2f, it's easy to find that, with this pre-processing, the typical segmentation method (based on CbCr Gaussian skin model or CbCr ellipse skin model) segmented fewer non-skin pixels and more real skin pixels, the owe or excessive segmentation was decreased.
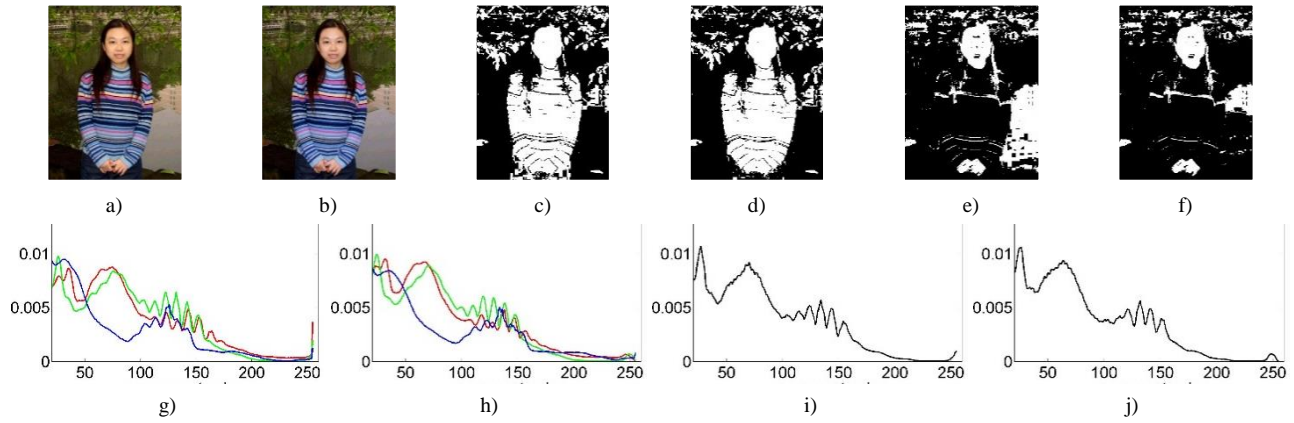
FIGURE 2 Image pre-processing: a) input image, b) pre-processed image, c) skin-segmented image of a) based on CbCr Gaussian model, d) skin-segmented image of b) based on CbCr Gaussian model, e) skin-segmented image of a) based on CbCr ellipse model, f) skin-segmented image of b) based on CbCr ellipse model, g) probability density distribution of R,G,B components of a), h) probability density distribution of R,G,B components of b), i) probability density distribution of Y component of a) j) probability density distribution of Y component of b)

4 Skin segmentation

Skin colour space and skin model used by colour image is very useful and important for skin segmentation and skin extraction. Used colour space commonly include RGB, YCbCr, YIQ, HSV, etc., the basic colour space is RGB, while the other colour spaces are transformed from RGB. The typical skin models include Gaussian model [1, 8], ellipse model [6], and the other models [9]. The single skin model may lose some information of skin colour and facial features in input image. Especially for image with colour excursion and high light, the insufficient and over segmentation based on a single skin model often occurred. The model's ability of skin description should be enhanced to improve accuracy of skin segmentation, and application scope of skin model still need to enlarge. To solve these problems, this paper built a CbgCbr-YIQ dual skin model to acquire the skin similarity, which reflected the distribution of skin colour better and get better segmentation results than a single skin model.
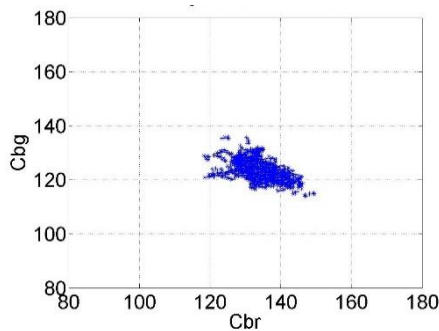
FIGURE 3 Skin color cluster in CbgCbr 2D color space

4.1 THE CBGCBR SKIN MODEL

The CbgCbr skin model originated from the CbgCbr 2D colour space and the average thought. As mentioned above, the pre-processed image (Figure 1b) can be acquired by pre-processing the original input image (Figure 1a). The original input image mainly came from the MIT single face test set [10] and self-built multi-face test set. This paper regarded the skin samples of pre-processed images as statistic sources. According to statistics, skin cluster analysis of skin samples showed that, the distribution of skin colour cluster presented a 2D ellipse in CbgCbr 2D colour space (Figure 3). So the CbgCbr ellipse skin model was built on the basis of skin cluster (Equations (6)-(9)). Based on CbgCbr ellipse skin model, the skin similarity can be calculated (Equation (9)), and skin-similarity image can be acquired (Equation (10) and Figure 1(c)).The steps of similarity calculation based on CbgCbr model are as follows:

Step 1: Acquire $Y'_{i,j}$, $Cr_{i,j}$, $Cg_{i,j}$, $Cb_{i,j}$ components of pre-processed image using Equations (6)-(7).

$$Y'_{i,j} = 0.299 \times R'_{i,j} + 0.578 \times G'_{i,j} + 0.114 \times B'_{i,j}, \qquad (6)$$

$$\begin{bmatrix} Cr_{i,j} \\ Cg_{i,j} \\ Cb_{i,j} \end{bmatrix} = \begin{bmatrix} 128 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.713 & 0 & 0 \\ 0 & 0.587 & 0 \\ 0 & 0 & 0.564 \end{bmatrix} \begin{bmatrix} R'_{i,j} - Y'_{i,j} \\ G'_{i,j} - Y'_{i,j} \\ B'_{i,j} - Y'_{i,j} \end{bmatrix}, (7)$$

Step 2: Use Equation (8) to calculate three average components: $Cbr_{i,j}$, $Cgr_{i,j}$, $Cbg_{i,j}$.

$$Cbr_{i,j} = \frac{Cb_{i,j} + Cr_{i,j}}{2}, Cgr_{i,j} = \frac{Cg_{i,j} + Cr_{i,j}}{2},$$

$$Cbg_{i,j} = \frac{Cb_{i,j} + Cg_{i,j}}{2} \qquad (8)$$

Step 3: Build the CbgCbr ellipse skin model. Based on this model, calculate the similarity by Equation (9). Acquire the similarity image by Equation (10) and the skin-segmented image by Equation (11).

$$S_{i,j} = \frac{(Cbr_{i,j} - 135)^2}{(12)^2} + \frac{(Cbg_{i,j} - 115)^2}{(10)^2}, \qquad (9)$$

$$P_{i,j} = \begin{cases} S_{i,j} \times 255 & if \left(S_{i,j} \leq 1 \right) \\ 255 & else \end{cases} \text{m}, \qquad (10)$$

$$F_{i,j} = \begin{cases} 255 & if \left(S_{i,j} \leq 1 \right) \\ 0 & else \end{cases} \text{m}, \qquad (11)$$

where $P_{i,j}$ and $F_{i,j}$ are the pixel's value of skin-similarity image and the pixel's value of skin-segmented image based on CbgCbr skin model.

4.2 THE ADAPTIVE YIQ SKIN MODEL

Lots of experiments and existing research demonstrated that the appearance of skin colour shows some degree clustering characteristic in YIQ colour space. The YIQ skin model originated from the clustering distribution of skin colour in the YIQ 2D colour space and the difference thought between I and Q components. This paper built an adaptive skin model in YIQ colour space to calculate skin similarity (Equation (13)) of pre-processed image, and correspondingly to acquire the similarity image (Equation (14) and Figure 1d). The steps of skin similarity calculation based on YIQ skin model are as follows:

Step 1: Acquire $Y'_{i,j}$, $I_{i,j}$ and $Q_{i,j}$ components of pre-processed image using Equation (12):

$$\begin{bmatrix} Y'_{i,j} \\ I_{i,j} \\ Q_{i,j} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.212 & -0.523 & -0.311 \end{bmatrix} \begin{bmatrix} B'_{i,j} \\ G'_{i,j} \\ R'_{i,j} \end{bmatrix}, \qquad (12)$$

Step 2: Build the adaptive YIQ skin model. Based on this model, calculate the similarity by Equation (13). Acquire the similarity image by Equation (14) and segmented image by Equation (15).

$$\begin{cases} t'_{i,j} = \frac{(I_{i,j})^2}{(Y'_{i,j})^2} + \frac{(Q_{i,j})^2}{(Y'_{i,j})^2}, \\ t''_{i,j} = \frac{\sqrt{(I_{i,j} - Q_{i,j})^2}}{Y'_{i,j}}, \quad S'_{i,j} = \frac{t'_{i,j}}{t''_{i,j}} \end{cases}, \qquad (13)$$

$$P'_{i,j} = \begin{cases} S'_{i,j} \times 255 & if \left(S'_{i,j} \leq 1 \text{ and } I_{i,j} > 20 \right) \\ 255 & else \end{cases}, \qquad (14)$$

$$F'_{i,j} = \begin{cases} 255 & if \left(S'_{i,j} \leq 1 \text{ and } I_{i,j} > 20 \right) \\ 0 & else \end{cases}, \qquad (15)$$

where, the $P'_{i,j}$ and $F'_{i,j}$ are the pixel's value of skin similarity image and the pixel's value of skin-segmented image based on YIQ skin model.

4.3 THE INTEGRATED SIMILARITY BASED ON THE DUAL SKIN MODEL

The calculation of the integrated similarity is to get the average value of two kinds of similarities based on the CbgCbr model and the YIQ adaptive model. It can combine advantages of two skin models, and make the integrated skin similarity had better accuracy, adaptability, and ability of skin description, especially in the case of different races, sophisticated background and variant light. So after acquiring two kinds of similarities ($S_{i,j}$ and $S'_{i,j}$), the integrated similarity $S''_{i,j}$ based on the dual skin model can be acquired using Equation (16). So the integrated similarity image is an image that fused two similarity images conditionally (Equation (17) and Figure 1e).

$$S''_{i,j} = \frac{(S'_{i,j} + S_{i,j})}{2}, \qquad (16)$$

$$P''_{i,j} = \begin{cases} S''_{i,j} \times 255 & if \left(S''_{i,j} < 1 \text{ and } S_{i,j} < 1 \text{ and } S'_{i,j} < 1 \right) \\ 255 & else \end{cases}, \qquad (17)$$

$$F''_{i,j} = \begin{cases} 255 & if \left(S''_{i,j} < 1 \text{ and } S_{i,j} < 1 \text{ and } S'_{i,j} < 1 \right) \\ 0 & else \end{cases}, \qquad (18)$$

$$\begin{cases} R''_{i,j} = R'_{i,j}, G''_{i,j} = G'_{i,j}, B''_{i,j} = B'_{i,j} \\ R''_{i,j} = G''_{i,j} = B''_{i,j} = 0 \text{ if } \left(S''_{i,j} \geq 1 \text{ or } S_{i,j} \geq 1 \text{ or } S'_{i,j} \geq 1 \right) \end{cases}. \qquad (19)$$

After acquiring the integrated skin similarity, by judging and binary segmenting non-white pixels of integrated similarity image (Equation (18)) and reserving corresponding region of pre-processed image (Equation (19)), skin-segmented image and skin-extracted image can be acquired (Figures 1f and 1g).

Figure 4 shows the examples of skin segmentation based on several skin models, including CbCr Gaussian model, CbCr ellipse model, CbgCbr ellipse model, YIQ adaptive model, CbgCbr-YIQ dual model. These models were built for the pre-processed image (Figure 4(a)). Comparing Figures 4d and 4e, or Figures 4j and 4k, note that the skin regions approximately have the same positions in two images, but the non-skin regions or noise regions are different partly. Based on these characteristics above, the integrated similarity image (Figure 4f) fused two similarity images by averaging and logical conditions

(Equations (16) and (17)) to preserve the real skin regions and remove the non-skin regions. From Figures 4j, 4k and 4l, it's easy to find that the skin-segmented image based on dual skin model had fewer non-skin pixels and preserved more real skin pixels, the owe segmentation or excessive

segmentation was decreased. Comparing Figure 4h, 4i and 4l, it is obvious that relative to the CbCr Gaussian model and CbCr ellipse model, the segmented image based on dual skin model had less error segmentation.
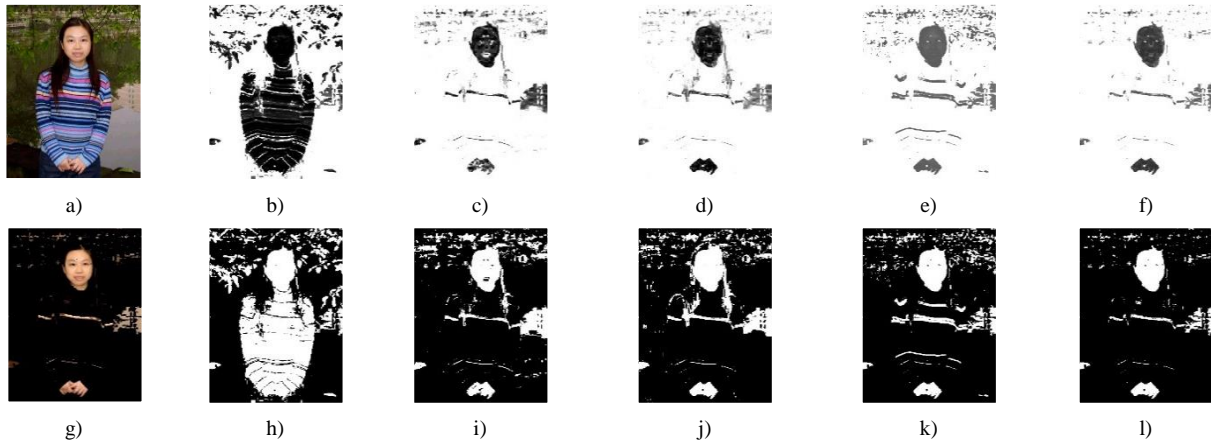
FIGURE 4 Skin segmentation and extraction: a) pre-processed image, b) similarity image of a) based on CbCr Gaussian model, c) similarity image of a) based on CbCr ellipse model, d) similarity image of a) based on CbgCbr ellipse model, e) similarity image of a) based on YIQ model, f) similarity image of a) based on CbgCbr-YIQ dual model, g) skin-extracted image based on f), h) skin-segmented image based on b), i) skin-segmented image based on c), j) skin-segmented image based on d), k) skin-segmented image based on e), l) skin-segmented image based on f)

5 Post-processing and localization

As stated above, the skin-segmented image and the skin-extracted image can be acquired based on the dual skin model. The skin-segmented image (Figure 4l) and the skin-extracted image both contained all the possible skin regions, including the real face objects, the non-face skin objects (such as hands, foots, legs, etc.), non-skin objects (skin-similar backgrounds), and some noise. In order to search the face regions and localize the facial features roughly as soon as possible, the morphological post-processing, the Adaboost classifier [24], the gradient [11] and difference technology were used in this part jointly. To depress or remove the noise interference, and to acquire

the face candidates, the post-processing adopted the morphological close operation and open operation, using the 3×3 template. With the close operation, the denoised skin-segmented image and the corresponding denoised skin-extracted image can be acquired (Figure 5b and 5h), which preserved the facial feature holes (e.g. eyes and mouth). With the open operation, the hole-filled skin-segmented image and the corresponding hole-filled skin-extracted image can be acquired (Figure 5c and 5i), which contained the all the face candidates. Moreover, the face candidates in hole-filled skin-extracted image (Figure 5i) are about to be detected or classified by trained Adaboost classifier.
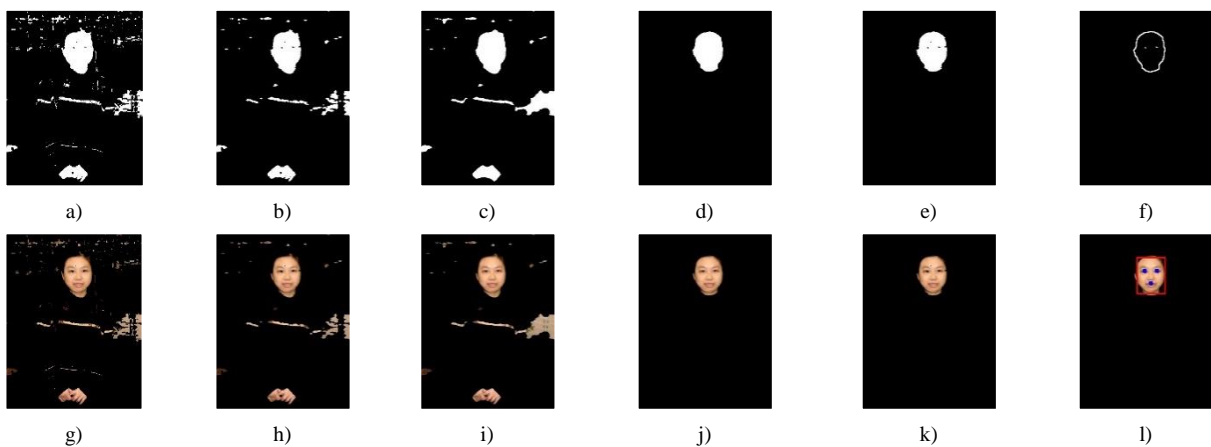
FIGURE 5 Post-processing and facial features localization: a) skin-segmented image based on CbgCbr-YIQ dual skin model, b) denoised skin-segmented image, c) hole-filled skin-segmented image, d) hole-filled face-segmented image, e) denoised face-segmented image (with facial feature holes), f) facial feature image g) skin-extracted image based on CbgCbr-YIQ dual skin model, h) denoised skin-extracted image, i) hole-filled skin-extracted image, j) hole-filled face-extracted image, k) denoised face-extracted image (with facial feature holes), l) face and facial feature localized image

As the hole-filled skin-segmented image and hole-filled skin-extracted image contained all the face candidates, some of which is the possible non-face objects. For removing the non-face regions, and acquiring the real face regions, the Adaboost classifier was used. There was a problem that how to narrow the range of searching or scanning. To solve this problem, the face candidates in the hole-filled skin-extracted image (Figure 5(i)) were regarded as the being detected or classified targets by Adaboost classifier, which was trained by face samples in the training set. Generally, the used object detected by classifier is a whole image, which spent too much time to scan or search all the pixels in the whole image. This classifier scanned the face candidate regions instead of all the pixels in the whole image. After judging and classifying, it discarded the non-face objects and preserved the face objects in hole-filled skin-extracted image (Figure 5(i)). So the extracted face target in the hole-filled face-extracted image can be acquired, as shown in Figure 5(j). This way not only can improve the speed of scanning and save the time of searching, but also increase the accuracy of detection. With Adaboost classifier, face regions in hole-filled face-extracted image can be got (Figure 5j), and the corresponding face-segmented image (Figure 5d) can be acquired. Moreover, the denoised face-segmented image with facial feature holes (Figure 5e) and corresponding denoised face-extracted image with holes (Figure 5k) also can be acquired. By comparison, it is thus clear that Figures 5d and 5e is different in facial feature regions (eyes and mouth regions). Using the difference technology and Gradient limitation, the facial feature image can be acquired (Figure 5f), and then facial feature localization can be realized (Figure 5l).

6 Experiment

The proposed method was tested on MIT face test set [10] and a self-built test set to evaluate its performance. The MIT face test set contains 50 images of 10 individuals, with different colour bias, variant light and poses. The self-built test set contains 100 images with 585 faces, gathered with different poses and expressions, variant light and sophisticated background. Each image can be rescale automatically to a standard size (150×150 pixels). The test condition is: P4, 2GHz CPU, 2G memory, and VC6. Based on two groups of test sets (single-face and multi-face), the average time in each stage of this algorithm was shown in Figure 6, where the average time on single-face test set was slightly higher than those on multi-face test set. Table 1 listed comparison of localization rates of proposed algorithm and the other algorithms. It is easy to find that the localization rate of this algorithm is higher than that of others, but the speed is fast. There are five examples shown in Figure 7, which contained the images with variant poses, variant light, sophisticated background, multiple faces.
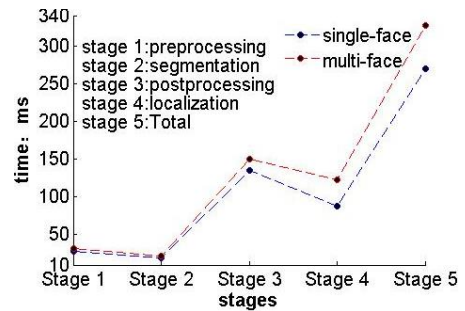
FIGURE 6 Average time in each stage

TABLE 1 Comparison of localization rate (Image Size:150×150)

No	Algorithm	Localization Rate	Test samples (No. of images/faces)
1	Algorithm based on template matching	80.8%	30 colour images (150)
2	Algorithm based on Gaussian model	84.1%	25 colour images (125)
3	Algorithm based on ellipse model	82.9%	25 colour images (125)
4	Algorithm based on Gaussian model+template matching [1]	97.2%	25 colour images (150)
5	Algorithm based on Bayes [3]	95.6%	20 grayscale images (130)
6	Algorithm based on Adaboost [4]	85.7%	30 colour images (150)
7	Algorithm based on Neural Network [5]	84.5%	25 grayscale images (125)
8	Algorithm based on lighting compensation+ ellipse mode [6]	92.5%	35 colour images (175)
9	Algorithm based on Gaussian model+PCA [12]	93.3%	25 colour images (150)
10	Proposed Algorithm in This Paper (the dual model+Adaboost)	98.3%	150 colour images (635)
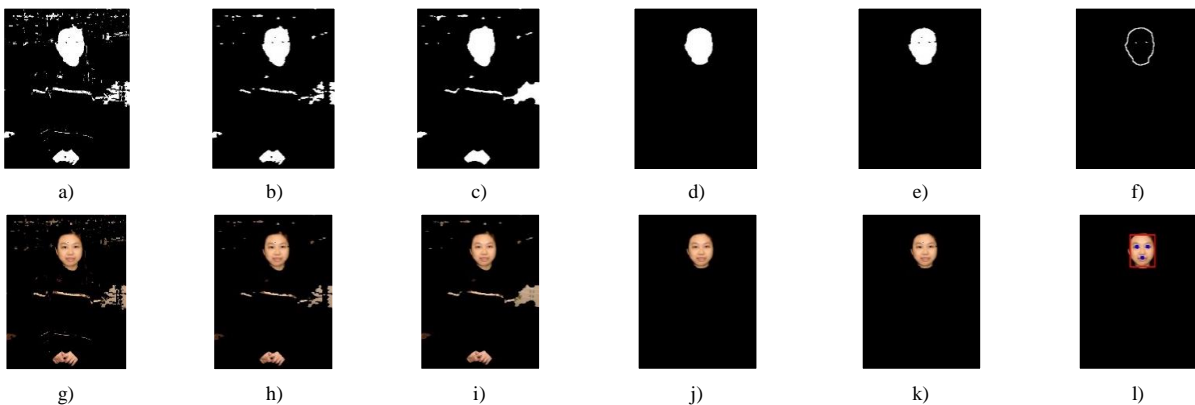
FIGURE 7 Examples of facial feature localization: a) d) g) j) m) p) colour input images, b) e) h) k) n) q) facial feature localized images (background removed), c) f) I) l) o) r) facial feature localized images

201

7 Conclusions

A Face and facial feature localization algorithm based on CbgCbr-YIQ dual skin model was presented in this paper. To improve the accuracy of skin segmentation, as well as the speed of the face and facial feature localization, the proposed algorithm contains pre-processing, building the dual skin model, skin segmentation and extraction, post-processing, face and facial feature positioning. The pre-processing depressed colour excursion and high light partly. Then the built dual skin model had better ability of skin description, adaptability and robustness than the single skin model. The skin segmentation and extraction based on dual skin model obtained fewer non-skin objects and more skin objects. The morphological post-processing removed noise fast, provided face candidates for Adaboost classifier and helped to decrease the range of scanning greatly. The face and facial feature positioning was realized by jointly using the classifier, the difference and gradient limitation technologies. The experimental results demonstrated that this algorithm had robustness and good performance for colour face image with sophisticated background, variant light, multiple poses and expressions.

References

[1] Wang Z, Li S 2011 *Information Technology Journal* **10**(12) 2308-14
[2] Guo J M, Lin C C, Wu M F, Chang C H, Lee H 2011 *IEEE Signal Proc Let* **18**(8) 447-50
[3] Wang Y, Wu Y 2010 *Pattern Recognition* **43**(3) 1008-15
[4] Yang M, Crenshaw J, Augustine B 2010 *Computer Visual and Image Understanding* **114**(11) 1116-25
[5] Wang Z, Li T 2008 A Face Detection System Based Skin Color and Neural Network *Proceedings of 2008 International Conference on Computer Science and Software Engineering* **1** 961-4
[6] Hsu R L, Mottaleb M A, Jain A K 2002 *IEEE Transactions on Patter Anal Mach Intell.* 24(5) 696-706
[7] Ibrahim N B, Selim M M, Zayed H H 2012 A dynamic skin detector based on face skin tone color *Proc of 2012 8th Int Conf on Informatics and Systems* MM-1-MM-5
[8] Tan W R, Chan C S, Yogarajah P, Condell J 2012 *IEEE Tranasactopm on Industrial Informatics* **8**(1) 138-47
[9] Chu H, Xie Z, Xu X, Zhou L, Liu Q 2011 Inspection and Recognition of Generalized Surface Defect for Precise Optical Elements *Information Technology Journal* **10**(7) 1395-401
[10] *CBCL Face Recognition Database* http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html 15 Nov 2013
[11] Niu Z H, Shan S G, Chen X L 2009 *IEEE Signal Processing Letters* **16**(10) 897-900
[12] Shih F Y, Cheng S X, Chuang C F, Wang P S P 2008 *Patt Recog Artif Intell* **22**(3) 515-34

Author

Wei Li, born in February, 1982, Sichuan, China

Current position, grades: lecturer Computer school, China West Normal University.
University studies: BSc degree in computer science and technology from China West Normal University (P.R.C.) in 2004. MSc degree in communication engineering from the Chongqing University (P.R.C.) in 2007.
Scientific interest: pattern recognition, signal processing, image processing, and embedded system.
Publications: 8 papers.

A novel multi-factor simulation algorithm about tactical network connectivity reliability

Xiaolei Sun, Ning Huang*, Jian Zhou, Yue Zhang

School of Reliability and Systems Engineering, Beihang University, 100191, Beijing, China

Abstract

Tactical network is a cooperative engagement of a collection of mobile nodes without the requirements of any centralized access points or existing infrastructures. Its connectivity is very crucial. Thus, the calculation of network's connectivity reliability has great significance to further research of tactical network. At present, the study of network is mainly through analytical methods. Those methods build abstract models alone without considering the transmission characteristics and terrain environment of a tactical network. Meanwhile, the existing simulation methods like Monte Carlo, is too abstract and simple. The OPNET simulation can solve the above problems better. In this paper, we analysed the related characteristics of tactical network and designed tactical network connectivity calculating algorithm based on the OPNET simulation. Through the algorithm, we take the transmission and environment features into consideration. As a result, we analysed the effects of wireless transmission characteristics, mobile models and terrain environment on network two-terminal reliability in specific cases.

Keywords: tactical network, OPNET simulation, two-terminal reliability, mobile networking, wireless networks

1 Introduction

As a future digitized battlefield support platform, tactical network completes both horizontal and vertical data transmission in tactical combat zone. It not only provides command and situational awareness information for all echelons, enhancing the commanders' and weapons platforms' ability of information resource sharing, but also meets the multi-level, multi-mode, and multi-cover communication needs of the future digital battlefield. Therefore, tactical network is under the spotlight in the military field.

Tactical network requires troops to ensure reliable connectivity among combat units, during a variety of complex tasks in the tactical environment. Thus, analyse connectivity reliability accurately is especially important for tactical network. In recent years, many scholars began to study the connectivity reliability of wireless and mobile network. Gabe Sierra and Lara Kauehak [1] defined several parameters characterizing the ad hoc networks, but there are no widely recognized definition and expression. Gabe Sierra [2] found that in the mobile wireless network there are still existing faults like large delay, high bit error rate and channel congestion, even when all the equipment are in good condition, so the reliability analysis method in the fixed topology network is difficult to be directly applied to the ad hoc network. Kharbash [3] took the mobility of nodes into consideration in his reliability assessment, but the evaluation model of wireless transmission characteristics was not given. [4, 5] used some mathematical models to evaluate the environment of wireless network, such as the Random Distribution Model

and the Grey Relational Model. But they couldn't give a complete modelling of the environmental factors in ad hoc network. The complex terrain and weather could not be considered in detail in their models. All in all, current analyses of network merely complete reliability modelling, rarely take many features of a mobile network into account [6, 7], such as node transmission characteristics, topography, terrain and other environmental factors. On the other hand, the simulation methods existed, like Monte Carlo, is always too abstract and simple. For example, Wang Xuewang [8] used the distance between nodes to replace other terrains factors. The OPNET simulation can solve the above problems, but it is lack of specific reliability analysis module.

In this paper, we propose a tactical network two-terminal reliability analytical method based on the OPNET simulation. This method is based on the OPNET modeller and includes three parts, i.e., the network modelling, the simulation data collecting and the connectivity reliability calculating. OPNET network simulation belongs to the "network inside simulation" [9]. Namely, the network is simulated by dynamic behaviour, and we can analyse the network characteristic combining the accessed performance data. What's more, we can import the authentic terrain data into the simulation, which filling the gaps mentioned aforesaid. On this basis, the paper brought out a kind of network two-terminal reliability experimental method and comparatively analysed networks in different mobile models and environments.

The rest of this paper is organized as follows: Section 2 shows the design of simulation algorithm and experiment. Section 3 shows experiment with random

---

Corresponding author e-mail: hn@buaa.edu.cn

Computer and Information Technologies

mobile module and without terrain. Section 4 shows the influence of different mobile modules. Section 5 analyses the influences of terrain. Finally, Section 6 states our conclusions.

2 Design of algorithm and experiment

We define the two-terminal reliability *R(G)*, is as following: the probability that the destination node can maintain connectivity with the source node, receiving packets sent by the source node successfully. In a tactical network $G = (V, E)$, $V=\{v1,v2,...vN\}$ is a set of *N* nodes in the network, and $E = \{e1，e2,...,eM\}$ is a set of *M* links in the network. There are only two statuses of the link between the source node *vi*, and the destination node *vj*, connected or not, i.e. *Ψ(X)=1 or Ψ(X)=0, Ψ(X)* is a Bernoulli random variable [10]. Assuming that *S* is the state space of *X*, and *X1,X2,...,XK* is the random *K* independent samples from *S*. Then the estimated value of the two-terminal reliability is as Equation (1).

$$R(G) = \frac{1}{K}\sum_{i=1}^{K}\psi(X)^i . \qquad (1)$$

Then we can get the value of *R(G)* through our multi-factor simulation algorithm, and use that value to analyse the reliability of the network. The process of the algorithm can be briefly described as follows:

Step 1. We build a project wireless-reliability-test in OPNET, and create several scenarios in it. Then, we put 30 wlan_wkstn_adv wireless mobiles nodes in one of the scenarios, and copy all of them to the other scenarios, ensuring that the nodes in all the scenarios have the same coordinates.

Step 2. Set the costumer business and routing protocol. We make the setting that source node sends destination node a 360 bytes data packet every 360 seconds, and choose the AODV routing protocol [9].

Step 3. Import the terrain data to the scenario if necessary. Import 100 random seeds into the project, and run the scenarios at the same time to get a set of reliability result from each scenario. Put the mobile module in all scenarios and set the required mobile model for all the nodes in two scenarios.

Step 4. Repeat the simulation for 50 different sets of random seeds. Average the results and draw the two-terminal curves for all the scenarios.

In this paper, we conducted three experiments. In all the three simulation experiments, we first establish a network scenario according to the parameters in Table 1. After that, we get random seeds form a C++ program, and obtain the K independent samples by importing the seeds into the network scenario in OPNET. At last, we can get the estimated value of the two-terminal reliability through Equation (1). The specific parameters of the simulation are shown in Table 1.

TABLE 1 Simulation experiment parameters

Parameters	Value
Network size	Scenario1000m², nodes 30
Simulation times	5000times (50*100)
Simulation length	100h
Transmission range	100m
Source node	Node 1
Destination node	Node 30
Mobile model	Random movement
Packet size	360byte
Packet sending interval	360s

3 Experiment with random mobile module and without terrain

We create a 1000 m*1000 m network in OPNET, and put 30 "wlan_wkstn_adv" wireless mobile nodes into the scenario randomly. Then we add task module, application module and profile module to the network scenario in order. The finished network scenario is shown in Figure 1.
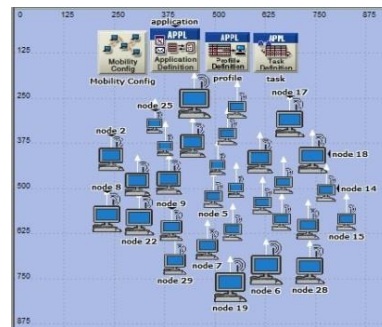
FIGURE 1 The first simulation experiment scenario

After building the network scenario, we begin to set the node parameters. Firstly, set node transmission power and reception sensitivity to 0.005W and - 81 respectively, so the maximum transmission distance will be 100 m according to the Equation (2).

$$Los = 32.44 + 20\lg d(km) + 20\lg f(MHz) . \qquad (2)$$

Then, we set Demand Distance Vector Routing (AODV) for all nodes, and add OPNET Default Random Waypoint Model to all the nodes in "Topology-Random mobility" [11]. We complete the custom business setting in the task, application and profile modules. Let Node 1 send a 360 bytes packet to Node 30 every 360 seconds. Under the circumstance that these two nodes are always connected, it can be seen that the destination node receives 1 byte data every second on average. Figure 2 shows the specific business setting.
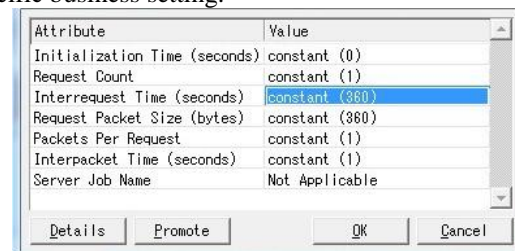
FIGURE 2 Custom business settings

Sun Xiaolei, Huang Ning, Zhou Jian, Zhang Yue

After finishing the setting process of node and business, right click on the scene and select Traffic Received (bytes/sec) as the statistics we need in Chose Individual DES Statistics. Set the simulation length to 100h and import the 100 random seeds generated in C++ program into the OPNET, and click run to start the simulation.

After running the process, we can get the results of byte data, which the source nodes received every second in 100 h, corresponding to the 100 random seeds. Based on the Equation (1), we can obtain a connectivity reliability estimation curve from the above 100 samples. Repeat this process 50 times and average the 50 results, then we get the final two-terminal reliability curve without terrain and random mobile model (Figure 3).
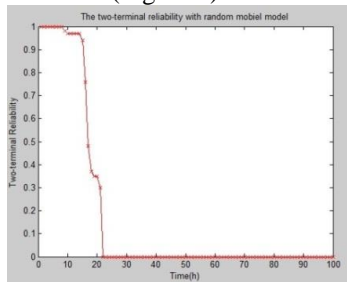
FIGURE 3 Final two-terminal reliability

In the 50 times of experiments, the most volatile point is at 17th hour. As showed in Figure 4. Through the calculation, we obtain the variance of the 50 reliability at 17th is 0.03, within acceptable limits. Therefore, the result of the simulation is quite stable.
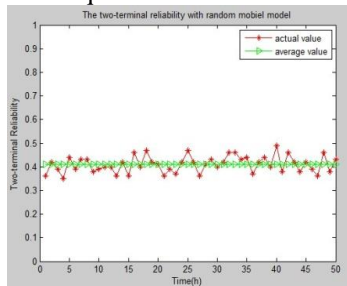
FIGURE 4 The stability test of the result

4 Experiment with different mobile modules

In the tactical network, the moving pattern of wireless nodes will directly affect the network topology, and affect the reliability of the network. In this chapter, we will study the different moving patterns' influence on two-terminal reliability according to the above simulation method [8]. In this paper, we use two-dimensional normal cloud model to create node mobile model. Two-dimensional normal cloud model can effectively describe the fuzziness and randomness of typical tactical moving pattern. See in Equation (3).

$$\mu_i(t) = \exp(-(\frac{(x_i(t) - E_x(t))^2}{2 E_{nx_i}(t)} + \frac{(y_i(t) - E_y(t))^2}{2 E_{ny_i}(t)})) . \qquad (3)$$

In Equation (3), (*Ex (t), Ey (t)*) is a two-dimensional normal random variable, whose expectation is (*Ex (T), Ey (T)*) and entropy is (*Hex2 (t), Hey2 (t)*). (*xi(t), yi(t)*) is also a two-dimensional normal random variable, whose expectation is *(Ex (t), Ey (t))* and entropy is *(Enxi2 (t), Enyi2 (t))*. *μi(t)* is a measured of *(xi(t), yi(t))* belonging to some degree domain. The two-dimensional normal cloud model could create five different mobile models depending on the different parameters selection of *(Ex, Ey, Enx, Eny, Hgx, Hgy, n)* [8]. Through some programming work, we develop the five normal cloud mobile modules in OPNET.

According to the algorithm method in Section 2, we add the five mobile models to the network scenario and run the simulation in proper order. After collecting the simulation data and calculating the two-terminal reliability, we can get the reliability comparison result curves of the five mobile models in Figure 5.
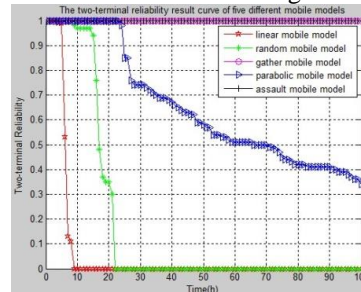
FIGURE 5 Connectivity reliability curve of the five mobile model

It can be seen from Figure 5 that, different from the traditional products fault, the two-terminal reliability between assigned source node and destination node in tactical network is not so significantly affected by time. But with different influencing factors, variation trends have large differences. Like the different tactical mobile model in this section, it has great influence on two-terminal reliability.

The reason is that two-terminal reliability mostly depends on the relative location of the assigned two-terminal nodes in the scenario and other nodes distribution situation between them, while the latter is greatly influenced by the mobile model. So the assigned moving pattern of nodes is significantly important to the two-terminal reliability of the network.

5 Experiment with terrain data

Terrain environment has an effect on the connection reliability of tactical network in several aspects. In this section, we study the impact on reliability caused by the terrain environment. Figure 6 shows the scenarios with terrain environment.
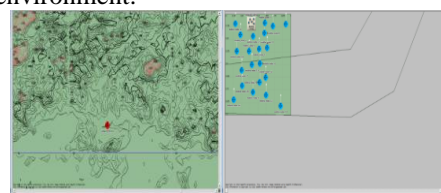
FIGURE 6 The scenarios with terrain environment

We set the parabolic mobile module to nodes in the above network, and through our algorithm, we get the curves both in scenarios with and without terrain environment. Figure 7 shows the details.
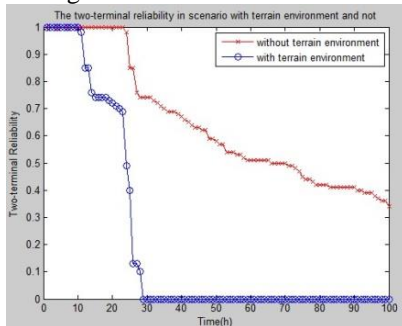
FIGURE 7 Reliability graph with and without terrain environment

It can be seen from Figure 7 that the terrain environment not only causes the value of two-terminal reliability decreases earlier, but also accelerates the decline rate. It is clear that the terrain environment dose have a significant effect on two-terminal reliability. The terrain environment can influence the connectivity reliability of the tactical network in several aspects. The two main reasons are given below:

1) With terrain environment under consideration, the max transmission distance decreases. Because signal attenuation between nodes is greater than that in a vacuum

environment, which is caused by terrain shelter, and then the connection reliability reduces at both ends.

2) Terrain factor has much effect on the moving speed and location of mobile nodes, which affect the topological structure of networks, and finally impact the two-terminal reliability.

6 Conclusions

This paper provides a detailed simulation analysis and computing method of the two-terminal reliability of tactical network based on OPNET, upon which we could finish experiments and analyse the effect that complex terrain environment and mobile models in tactical network have on connectivity reliability. The innovative point of this paper is that it finished the analysis of the effect terrain environment and mobile models have on connectivity reliability, which is very difficult to be analysed comprehensively by traditional analytical methods.

In the future, we will apply our analysis method to the study of other network connectivity reliability aside from the two-terminal reliability. What's more, we want to expand use of this method to the research of network performance reliability. For these purposes, we will perform some other practical cases for the utilization of the proposed analytical method further.

References

[1] Kharbash A 2007 Computing two-terminal reliability in mobile ad hoc networks *WCNC* 2831-6
[2] Gabe S, Lara K 2006 Reliability of the wireless ad hoc network *www.CEC- hinamag.com*
[3] Geir E 2009 The availability and reliability of wireless multi-hop networks with stochastic link failures *IEEE Journal on Selected Areas in Communications* **27**(7) 1132- 46
[4] Cho J, H, Swami A, Chen I R 2012 Modeling and analysis of trust management with trust chain optimization in mobile ad hoc networks *Journal of Network and Computer Applications* **35**(3) 1001-12
[5] Joe M, Henri R, Alan L, Rob Y 2006 object oriented design in real-time hardware-in-the-loop simulation *Proc Of SPIE* V6208 680806 1-10
[6] Eric B, Mike E, Craig M, Ranga R 2005 Network in the loop using HLA, distributed OPNET simulations and 3D visualizations *IEEE Military Communications Conference MILCOM 2005* 1-5

[7] Jiaxi C 2012 Research on network fault model and its influence based on OPNET *M.S. thesis, Dept Electron Eng Beijing University of Aeronautics and Astronautics Beijing China (in Chinese)*
[8] Xuewang W 2012 Research on connectivity and volume connectivity of tactical network *PhD dissertation Dept Electron Eng Beijing University of Aeronautics and Astronautics Beijing China*
[9] Nan F, Wei B, Xin L 2012 Mobile node position research and management module development *Beijing University of Aeronautics and Astronautics Beijing China*
[10] Sen L, Jie W, Bowen X, Xiaoguang W 2011 Monte Carlo simulation combining network reduction with the same network reliability *Computer Engineering* **37**(8) 1000-3428
[11] Nan F, Wei B, Xin L 2012 Research and development of network reliability based on simulation *Beijing University of Aeronautics and Astronautics Beijing China*

Authors

Xiaolei Sun, born on January 11, 1990, Zhejiang, China

Current position: master candidate in School of Reliability and System Engineering, Beihang University, second-year graduate student.
University studies: industrial engineering in Beihang University.
Scientific interest: network reliability. AD HOC network.
Publications: 1 patents, 2 papers.

Ning Huang, born on May 19, 1968, Kunming, China

Current position: Doctor of computer software, professor and doctoral supervisor in Beihang University, Chine.
University studies: computer science in Beihang University.
Scientific interest: network reliability, software test and software reliability.
Publications: more than 50 academic papers.

Jian Zhou, born on September 13, 1988, Hanzhong, China

Current position: PHD Candidate in Beihang University, Chine.
University studies: industrial engineering in Beihang University.
Scientific interest: network reliability, complex network and cascading failure.
Publications: 1 patents, 3 papers.

Yue Zhang, born on April 11, 1990, Changzhi, China

Current position: PHD Candidate in Beihang University.
University studies: industrial engineering in Beihang University.
Scientific interest: network reliability, network traffic modelling.
Publications: 1 patents, 2 papers

Video target tracking with fisher discriminant dictionary learning

Jian-Feng Zheng, Ji Zhang*

School of Information Science and Engineering, ChangZhou University, ChangZhou 213164, China

Abstract

As one of the state-of-the-art tracking methods based on sparse coding, l1-tracker finds the target with the minimum reconstruction error from the target template subspace. But the high computational costs restrict its application in practical terms heavily. In this paper, we incorporate the discriminant information into original l1-tracker, and introduce it into the tracking framework, called FD2LT. In our framework, tracking is considered as a problem consisting of object location with dictionary learned in the last frame in generative tracking framework, training samples selection, and dictionary learning with fisher discriminant dictionary learning (FDDL). With our method, the dictionary is much smaller than that in original one, moreover, without loss of tracking performance (and even better in some scenarios). The discriminant power explored from the dictionary is used in generative tracking. Experimental results demonstrate the effectiveness and efficiency of the proposed tracking algorithm.

Keywords: Fisher discrimination dictionary learning, generative and discriminant tracking, sparse coding, video object tracking

1 Introduction

Recently, computer vision is widely used in many fields. As one of the most exciting fields, target tracking looks for some specified objects pre-defined in video streams artificially. Targets change dynamically and uncertainly in video sequence, because of occlusion, noisy, varying and so on. Many tracking algorithms have been proposed, such as *IVT*, *TLD*, *CovTrack*, *l1-tracker* [1, 2].

Based on *sparse coding* (*SC*) [3], Mei proposed *l1*-tracker [4], where many challenging problems presented in tracking are addressed seamlessly. However, computational cost of *l1-tracker* is quite expensive to achieve efficient tracking. Moreover, the discriminant ability of the dictionary is not explored. An alternative way is to construct the dictionary with rich representation ability and few atoms, which is called *dictionary learning* (*DL*) [5]. Many *DL* algorithms have been proposed in last several years. *K-SVD* [6] learns the dictionary from training sets, which is suitable for reconstruction, rather than discrimination. Mairal introduces discriminant constraint into *K-SVD* for classification [7], which is not convex; Tosic proposes a new method for learning the over-complete dictionary to represent the stereo images [8], but not for classification like *K-SVD*; Yang's *Fisher discriminant dictionary learning* (*FDDL*) aims to learn a structured dictionary for face recognition, whose sub-dictionaries have specific class labels[5]. Our method is motivated by Yang's *FDDL*. The rest of this paper is organized as follows: sparse coding, *l1-tracker* and *FDDL* are introduced in section 2. In section 3, we analysis the shortcoming of *FDDL*, then improves and introduce it into tracking, called *FD2LT*. The convergence of *FD2LT* is demonstrated numerically. Experimental results with

FD2LT and some competitive algorithms are reported in section 4. Finally, we will conclude our work and propose future work.

2 Related Work

2.1 SPARSE CODING FORMULATION AND SPARSE CODING BASED TRACKING (*L1*-TRACKER)

SC is an attractive signal reconstruction method, and the main task of which is to reconstruct a query signal $y \in \mathbb{R}^{d \times 1}$ over the over-complete dictionary $D \in \mathbb{R}^{d \times n}$ with a sparse coefficient vector $x \in \mathbb{R}^{n \times 1}$:

$$\min_x \|y - Dx\|_F^2 + \lambda \|x\|_1. \tag{1}$$

where, $\|.\|_F$ and $\|.\|_1$ are the Frobenius-norm and l_1-norm, respectively. l_1-tracker is proposed based on *SC* [4], as shown in Figure 1. Suppose that, the target for tracking has been located in #205 (where the red box indicated and l_1-tracker initialized in #206), and *N* candidate regions are generated with Bayesian inference around it. With *n* templates learned from the last frame and 2*d* trivial templates (*d* positive ones and *d* negative ones, *d* is the dimension of 1-D stretched image) in Figure 1b), Equation (1) can be solved like Figure 1c). Furthermore, with these trivial templates, Mei adds *non-negative constraint* (*NNC*) *x≥0* into Equation (1). Reconstruction errors of all candidates with *SC* coefficients can be used to determine the weights for each candidate, the target in #206 can be located with the sum of weighted candidates, and the updating strategies of dictionaries can be seen in [4].

Corresponding author e-mail: zhangji@cczu.edu.cn

a) Candidate regions

b) Dictionary for l_1-optimization

c) SC of Candidates over dictionary
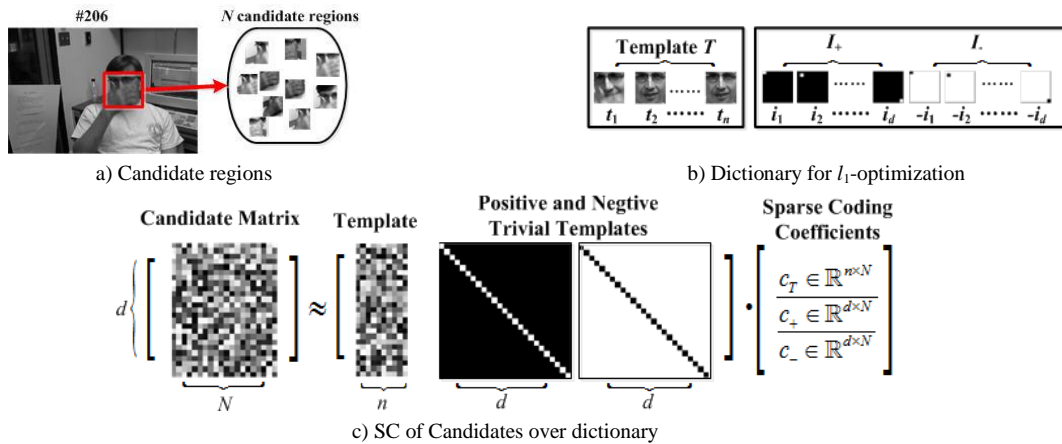
FIGURE 1 Original l_1-tracker

2.2 FISHER DISCRIMINANT DICTIONARY LEARNING (*FDDL*)

Mei's experiments show that, l_1-*tracker* has excellent performance when comparing with some state-of-the-art trackers [4]. But it is inefficiency, caused by the number of candidates (particles) N and the size of over-complete dictionary D, affords its application in real-time tracking severely. In original l_1-*tracker*, in order to achieve robust tracking, N must be very large, while the dimension of dictionary D is $d \times (2d+n)$ in Equation (1). In our experiment setting, N=200, and D for l_1-*tracker* is 1600×3210. It is quite nature that, how to reduce the number of candidates and the size of dictionary without (or with a little) loss of tracking accuracy, are two important issues in l_1-*tracker*. The former depends on the improvement of PF tracking framework [1], which is not mentioned in this paper; and the latter, specifically, how to construct dictionary which not only contains few atoms, but also has good ability of representation, is exactly the main task of our algorithm.

FDDL is proposed for face recognition, which learns c structured dictionaries $D=[D_1,D_2,\ldots, D_c]$ for each class of facial images, instead of a whole shared dictionary for all images, where D_i is the class-specified sub-dictionary associated with class-i, and c is the class number. Let $Y=[Y_1,Y_2,\ldots, Y_c]$ and $X=[X_1,X_2,\ldots,X_c]$ denote the set of training samples and the coding coefficient matrix of Y over D, respectively, where Y_i is the sub-set of the training samples from class i, X_i is the sub-matrix containing the coding coefficients of Y_i over D and $Y \approx DX$. *FDDL* can be formulated as following:

$$J_{(D,X)} = \min_{(D,X)} \left\{ r(Y,D,X) + \lambda_1 \|X\|_1 + \lambda_2 f(X) \right\}, \quad (2)$$

where, $f(X)=tr(S_W(X))-tr(S_B(X))-\eta\|X\|_F^2$ is a discriminative constraint imposed on X, which makes D discriminative for the samples in Y; $S_W(X)$ and $S_B(X)$ are within- and between-class scatters of X, respectively; λ_1, λ_2 are used to tune the influences of each term; $r(Y,D,X)=\sum_{i=1,\ldots c} r(Y_i,D,X_i)$ is the discriminative fidelity term, and:

$$r(Y_i, D, X_i) = \|Y_i - DX_i\|_F^2 + \|Y_i - D_i X_i^i\|_F^2 + \sum_{j=1, j\neq i}^{c} \|D_j X_i^j\|_F^2, \quad (3)$$

The first two terms ensure that Y_i can be represented by D and D_i approximately with X_i and X_i^i, respectively; the last one ensures the representation of Y_i over $D_j (i \neq j)$ is small. Some important terms in $r(Y_i,D,X_i)$ is shown in Figure 2. Consider some $y \in \mathbb{R}^{d \times 1}$ (e.g. a stretched face image), $\tilde{y} = Dx$ and $\hat{y} = D_i x_i$ are reconstruction results of y over the whole dictionary D and the class-i dictionary D_i, respectively. We denote the first two terms in Equation (3) as $\tilde{e}$ and $\hat{e}$ in Figure 2. The minimization of Equation (3) can be divided into two sub-problems: updating X by fixing D, and updating D by fixing X [9].

3 Our tracking framework with FDDL

3.1 IMPROVED FDDL

As mentioned above, Equation (3) minimizes $\tilde{e}, \hat{e}$ and $\sum_i y_j$ ($i \neq j$, $j=1,\ldots,c$) in Figure 2. But we find that, it is not sufficient for reconstructing signal y based on Equation (3). Denote by $y'=D_i x'$ the approximation of $\tilde{y}$ over D_i, and $e'=\hat{y} - y'$, $e = y - y'$, $e^*=\tilde{y} - y'$. Here, we use *AR* face database to validate the insufficient. *AR* contains 700 face images from 100 individuals (7 images for each one). In our experiment, 100 images are selected as query singles randomly, and the rest 600 images are treated as dictionary atoms. For each selected query image, $\|e\|_F$, $\|e^*\|_F$, $\|e'\|_F$, $\|\tilde{e}\|_F$ and $\|\hat{e}\|_F$ are calculated on 600 labelled training images, and plotted in Figure 3. It is clear to see that:

1) Minimization of $\|\tilde{e}\|_F$ and $\|\hat{e}\|_F$ cannot guarantee minimization of $\|e\|_F$, $\|e^*\|_F$ and $\|e'\|_F$;

2) As $\|e'\|_F > 0$ (unless the total dictionary D is consisting of ith class dictionary D_i merely, which is not practical), $\|\hat{e}\|_F < \|e\|_F$. And minimization of $\|\tilde{e}\|_F$ and $\|\hat{e}\|_F$ in Equation (3) has nothing to do with minimization of $\|e^*\|_F$;

3) Beside $\|\tilde{e}\|_F$ and $\|\hat{e}\|$, $\|e\|_F$, $\|e^*\|_F$ and $\|e'\|_F$ also

play important roles in face recognition. They report the difference between i^{th} class information contained in D_i and that in D. The smaller these three terms, the more positive information contained in D_i and the less in $D\backslash D_i$;

4) For each image, $\|e\|_F$ is maximum among all five residual terms, and the minimization of $\|e\|_F$ can be considered as the upper bound of all five representation residual terms. The similar results can be obtained on *Yale*, *ORL* database. Therefore, we can rewrite Equation (3) as:

$$r'\left(Y,D,X\right) = \sum_{i=1}^{c} r'\left(Y_i,D,X_i\right) = \sum_{i=1}^{c}\left(\left\|Y_i - D_i X_i'^{i}\right\|_F^2 + \sum_{j=1,j\neq i}^{c}\left\|D_j X_i'^{j}\right\|_F^2\right), \quad (4)$$

where, Y_i is the class-i subset of the training samples, X'_i^j is the coding coefficient matrix of $\tilde{Y}_i$ (reconstruction of Y_i over D_j). Note that, we use X'_i^j in Equation (4) instead of X_i^j in Equation (3). Denote $f(X')=tr(S_W(X'))-tr(S_B(X'))-\eta\|X'\|_F^2$, thus Equation (2) can be rewritten as:

$$J_{(D,X')} = \min_{(D,X')}\left\{r'\left(Y,D,X'\right) + \lambda_1\|X'\|_1 + \lambda_2 f\left(X'\right)\right\}. \quad (5)$$
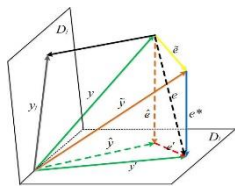
FIGURE 2 Residual terms   FIGURE 3 Residual terms
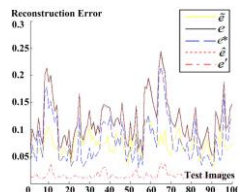
(a)Target in #204   (b)Candidates   (c)Sampling range   (d)Training set
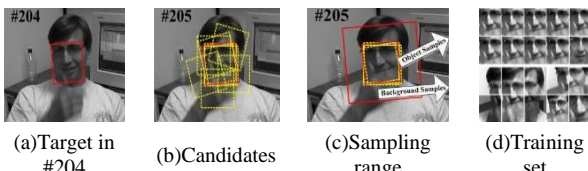
FIGURE 4 Object location and training samples selection

3.2 FDDL BASED TRACKING

In this subsection, our tracking framework called *FDDL based tracking* (*FD²LT*) is proposed, which includes three components: target location with the dictionary learned in the last frame, training samples selection for dictionary learning, and discriminant dictionary learning.

3.2.1 Target Location and Training Samples Selection

According to the target location in the last frame (red box in Figure 4a), a number of candidate regions can be extracted with Bayesian inference (dotted boxes in Figure 4b), then target can be distinguished from those candidates. Many *SC* methods have been proposed for classification [5-7]. Most of them (no matter supervised,

unsupervised and semi-supervised classification) work well based on a huge number of training samples, but as discussed above, *DL* with so many samples (e.g. *Y* in Equation (2-5)) lead to the tremendous computational costs. It is not critical for classification, as training and updating of classifiers are off-line in advance and classifying the new-arrival sample is very fast with the classifiers; but it is impatient in tracking, as the latter is real-time.

In order to achieve robustness and efficiency, under the framework of PF framework, we seek the most likely candidate region as target in current frame, then generate object/positive samples and background/negative samples as following. Select 10 regions extremely nearby the object (small red rectangle in Figure 4c) as positive samples Y_O, and 10 regions (including up-left, up, up-right, left, right, left-down, down, left-right, 1/3 bigger and 1/3 smaller than the small red box) as negative samples Y_B. Notice that, as shown in Figure 4d), in order to remain information of target during tracking, we always fix the last one in Y_O with the target selected artificially in the first frame; and the last two background regions are used to deal with the scale changing of target.

3.2.2 Dictionary Learning

Most of discriminant tracking methods are based on the assumption that, the appearances of target and background (near the object) change slightly frame by frame. Thus, we can represent candidate regions selected in current frame using the dictionary D_{old} learned in the last frame, and locate target as shown in the last section; afterwards, update D_{old} with 20 selected samples to help tracking in the successive frame. So, updating the dictionary is also a critical problem here.

Suppose that, $D_{old}=[D_{old\_O}, D_{old\_B}]$ is the last learned dictionary, and labeled training samples $Y=[Y_O,Y_B]$ are selected as Figure 4d). The problem of *DL* is how to update D_{old}, such that Y can be represented by new dictionary D_{new} with as less error as possible. According to improved *FDDL* proposed above, objective function for tracking is:

$$J_{(D',X')} = \min_{(D',X')}\left\{\left\|Y_O - D'_O X_O'^{O}\right\|_F^2 + \left\|Y_B - D'_B X_B'^{B}\right\|_F^2 + \left\|D'_B X_O'^{B}\right\|_F^2 + \left\|D'_O X_B'^{O}\right\|_F^2 + \lambda_1\|X'\|_1 + \lambda_2\left(S_W\left(X'\right) - S_B\left(X'\right) - \eta\|X'\|_F^2\right)\right\} \quad (6)$$

where, $\tilde{Y} = \left[\tilde{Y}_O, \tilde{Y}_B\right]$ are approximation of $Y = \left[Y_O, Y_B\right]$ over D_{old}; $X' = \left[X'_O, X'_B\right]$; $X'_O = \left[X_O'^{O}, X_O'^{B}\right]$ and $X'_B = \left[X_B'^{O}, X_B'^{B}\right]$ are SC coefficients of $\tilde{Y}_O$ and $\tilde{Y}_B$ over updated D', respectively.

3.2.3 Solving modified FDDL

Optimization of $J(D',X')$ in Equation (6) is not convex with respect to D' and X' simultaneously, and we can divided it into two sub-problems: updating X' by fixing D'; and updating D' by fixing X'. The procedures are iteratively implemented in our previous work, the reader can refer to [10] for detail.

3.3 REPRESENTATIVE/DISCRIMINANT ABILITIES OF DICTIONARY

Figure 5 shows some tracking results with our modified and original FD^2LT in #205 (Figures 5a and 5b) and #355 (Figure 5c) of *Dudek* sequence, and the further and detail results can be seen in Figure 8b). Experimental settings are the same as those in section 4.1. It is clear to see that:

1) Tracking results of our two FD^2LT methods are similar in #205, both of them deal with deformations and occlusions of target steadily;

2) FD^2LT with our improved *FDDL* remains object information well after dictionary updating, while the latter destroys almost all object information, which is shown in 10 object atoms (updated dictionary in current frame) in the lower-left corner of Figures 5a and 5b, respectively;

(3) We also assert that, coding coefficients used to represent target in #205 with improved FD^2LT is sparser than that with original FD^2LT. The button row of Figures 5a and 5b shows the coding results of these two methods.

With experiments in section 4, we get similar conclusions and list as following:

1) Our improved *FDDL* is much sparser than original *FDDL*, when coding the same signal;

2) Dictionary learning with our improved *FDDL* has stronger representation ability than original *FDDL*.

The second conclusion can be expressed in Figure 5c. With our method, after a few successive frames, some atoms in dictionary are almost close to zeros. As shown in Figure 5c, only 3 object atoms and 10 background atoms are used to represent the object with our improved *FDDL*, while it appears rarely when we use original *FDDL* in tracking. But it does not always benefit for tracking, especially when object changes heavily, as if the number of dictionary atoms is too small, they cannot remain the object information and adapt the change of object, simultaneously. In order to maintain rich ability of representation, we set those atoms with all zero elements as mean of all non-zero atoms.

(a) Original FD^2LT (b) Improved FD^2LT (c) Special Example
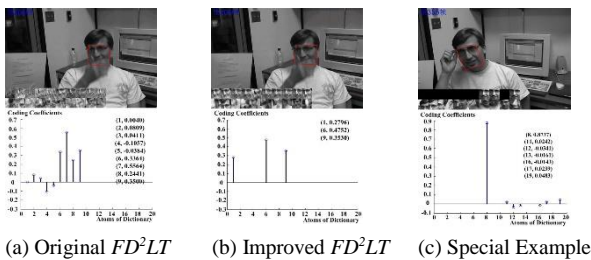
FIGURE 5 Results of FD^2LT on Dudek

FIGURE 6 NNC. Top: Tracking results of successive five frames with 1: l_1-tracker(Bule), 2: original FD^2LT (Red) and 3: improved FD^2LT (Yellow); Bottom: dictionary templates of three algorithms

3.4 WITH OR WITHOUT NON-NEGATIVE CONSTRAINTS?

In original l_1-*tracker*, *non-negative constraints* (*NNC*) for coding coefficients are not only used to restrict them to be positive, but also used to filter out clutter that is similar to target templates as reversed intensity patterns, as shown in Figure 6. In these successive five frames in *Car4* sequence, the car for tracking is driven from brightness to darkness. With *NNC*, original l_1-*tracker* remains the template information of the car, no matter it is in the shadow or not. But we have referred that, l_1-*tracker* costs heavily, because of the large dictionary(almost half of the dictionary, i.e. the negative trivial template, is designed for *NNC*) and l_1-*optimization*, as shown in Figure 1 and section 2.2, respectively.

In our FD^2LT, which can be considered as the discriminant extended version of l_1-*tracker*, *NNC* is omitted, because the changes of reversed intensity patterns can be incorporated into dictionary after learning. We find that, in Figure 1, the dictionaries learned with original and improved *FDDL* include not only positive templates, but also negative ones, which are used to deal with the problem of reversed intensity patterns. According to this, our two FD^2LT algorithms are much faster than original l_1-*tracker*, and the performances of them are nearly the same, which can be seen in Section 4.

To compare the convergence we verify that for our two FD^2LT trackers with *Dudek* sequence quantification ally. In #206, almost the entity face of the man (target for tracking) is blocked by his hand, which is slightly occurred in #205, see Figures 1a and Figures 5a, respectively. This is still a heavy representative change for modern trackers. There is reason to regard that, dictionaries learned in #205 with above two *FDDL* methods are not very suitable for tracking the object in #206. We get the same result in Figure 7a that, initial energies (values of objective function, calculated with Equation (7) are very large. During the iterative procedure, energies decayed quickly, and FD^2LT based on our improved *FDDL* convergences faster than the other one. Figure 7b shows the same result with the average energies convergence of the whole tracking procedure. In our experiment, we set number of iterations as 5, in order to keep our tracker working fast.
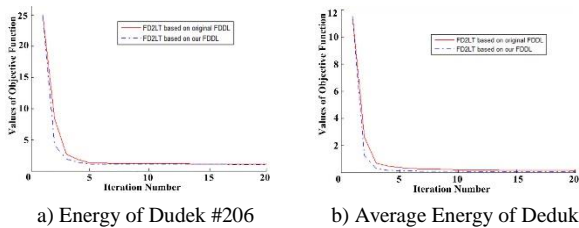
a) Energy of Dudek #206   b) Average Energy of Deduk

FIGURE 7 Convergence Curves of Energies (Values of Objective Function)

4 Experiment

4.1 EXPERIMENT SETTING

In order to evaluate our trackers, we conduct experiments on eight challenging video sequences, including *Surfer*, *Dudek*, *Faceocc2*, *Animal*, *Girl*, *Car11*, *Stone*, *Jumping*, *Car4* and *Pktest02* with 375, 1145, 819,71,500,393,200,100,400 and 120 frames, respectively (see Figure 8). These sequences cover almost all challenges in tracking, including occlusion (even heavy), motion blur, rotation, scale variation and complex background, etc. For comparison, we select four state-of-the-art trackers, including *incremental learning based tracker* (*IVT*, a familiar discriminant tracking method [11]), *Tracking -Learning-Detection* (*TLD*, a real-time long-term discriminant tracking method [12]), *CovTrack*(a generative tracker on Lie-group [13]) and l_1-*tracker*(a generative tracking methods[4])[1].

4.2 EXPERIMENTAL RESULTS

We evaluate the above-mentioned algorithms using the centre location errors, as shown in Figure 8. Overall, our FD^2LT performs well against the other state-of-the-art algorithms.

For occlusion, five algorithms except *IVT* work steadily roughly, especially at #206, #366 of the *Dudek* sequence in Figure 8b (head for tracking is covered by hand and glasses) and at #85, #108, #433 of the *Girl* sequence in Figure 8e (head for tracking turns right, turns back and blocks by someone else). After the target recovers from occlusion, these five trackers can seek it quickly. *IVT* works poorly, even lost the target from #10 in the *Girl* sequence, because of the number of positive and negative samples are limited (in consideration of the learning efficiency). Incremental updating of classifier in *IVT* is less effective; *CovTrack* has large size of candidates (with the definition of integral image, feature extraction of these candidates is so fast, and the costs of which can be ignored), which makes it robust for occlusion, scale variation and blur. Thanks to *P-N* expert learning and detection when loses the target, *TLD* often performs good when confronts occlusion. When occlusion happens, *TLD* abandons tracking, see the yellow regions in Figure 8.

When target appears again, *TLD* can obtain it. But there are also some exceptions, see Figure 8c from #377 to the end of the *Faceocc2* sequence. *TLD* loses the head for tracking from #377 because of the occlusion and rotation of the target, and after that, *TLD* never detects the target again; while original l_1-*tracker* and the derivative two FD^2LT trackers, which have strong representative abilities based on the large size of dictionary and good performances of dictionary learning, respectively, performance excellently.

For motion blur, our two trackers work better than *IVT* and l_1-*tracker*, moreover, *CovTrack* and *TLD* also reveal their abilities for blur, see #4, #9 and #38 in Figure 8d and #16, #29 and #53 in Figure 8h. The animal runs and jumps fast(motion blur) with splashing a lot of water splashing (occlusion). *IVT* and l_1-*tracker* fail both from #4, and never recover after that. Original and improved FD^2LT lost target at #28 and recover at #38, Figure 8d. And at #12,#21 and #43,#71, improved FD^2LT works better than original FD^2LT, *CovTrack* and *TLD*. *TLD* loses target from #24 to #33, from #53 to #71 in *Animal* sequence and from #33 to #36, from #41 to #48, from #56 to #70, from #73 to #90 in *Jumping* sequence.

For rotation and scale variation, our trackers still work robustly, see Figures 8a, 8c and 8e. The surfer staggers forward and back in the *Surfer* sequence, the girl turns left, turns right, zoom in and zoom out in the *Girl* sequence, and the man turns left, turns right and occludes by book in the *Faceocc2* sequence, four trackers except *IVT* and *TLD* perform nice for these challenges. Especially, *TLD* loses the target in #377 in *Faceocc2*, and never recovers again.

For complex background, as shown in Figures 8f and 8g the car for tracking is driven in the dark with bright lamplight and car light affecting the tracking, and the stone for tracking scatters around lots of similar stones. *TLD* and our two trackers work well before #220 in *Car11* sequence but *IVT* loses the target from #50. And in the *Stone* sequence, *TLD*, l_1-*tracker* and our two trackers work better than other two trackers as before. *CovTrack* fail in these two sequences, because it extracts edge information of targets as one dimension of features, and in these two sequences, edge of targets are ambiguous and hard to distinct. l_1-*tracker* fails after #220 in *Car11*, because of it is short of the discriminant ability of foreground and background; On the other hand, when there exists a candidate region which is like the target for tracking, it is likely for *TLD* to detect the former instead of the latter, as shown in #50 in Figure 8d and from #85 to the end in Figure 8j. This is because of the excessive strong detective ability of *TLD*, when losing the target.

In general, from above analysis, we can find that, owing to powerful representative and discriminant capabilities, our original and improved FD^2LT trackers work nearly the same, and the latter is slightly better, especially in the *Faceocc2*, *Dudek*, *Girl*, *Car11*, *Stone* and

---

[1] Readers can download codes of *IVT*(Matlab version) and *TLD*(C++ version) on www.cs.toronto.edu/ dross/ivt/ and info.ee.surrey.ac.uk/Personal/Z.Kalal/,

respectively. The other programs are coded with Matlab7.0 ourselves, and experiments are running on computer with 2.67GHz CPU and 2GB memory.

211

Car4 sequences, see Figures 8c, 8b, 8e, 8f, 8g, 8i; Original l_1-*tracker*[4] performances good in most frames, but fail to track sometimes; *TLD* has high performance in most of situations, which is worthy of "*long term tracker*"[12]. But it gives up tracking when facing heavily occlusion and rotation, and cannot recover when target appears again with large changes in appearance; *CovTrack* is suitable for occlusion and rotation, but fails when facing with complex backgrounds; *IVT* is sensitive, when the occlusion, rotation, motion blur of target are appeared in tracking. We also find that, our improved FD^2LT remains object information in updated dictionary, which are shown in button-two rows of lower-left of each frame in Figures 8a and 8b, while original FD^2LT destroys almost all object information, which are shown in top-two rows of lower-left of each frames. The same conclusions can be obtained when investigating the rest sequences in Figure 8. Moreover, the fixed 10th object training samples also prevent the procedure of *DL* from degeneration. See #28 in Figure 8d, all six trackers lost the target, and most of the dictionary atoms after updating (used to track in the successive frame) are confused, except few atom. With our selection of training samples for *DL* in section 3.1, our two trackers retrieve the animal's head in #38, but the other two methods fail. Meanwhile, *CovTrack* also preserves the original information of target, so it recovers in #34; and *TLD* uses detection module(not tracking or learning module) to search for the object and regain it in #38. All these mean that, remaining the original information of targets, instead of updating the whole templates frame by frame, is beneficial for tracking, which is taken more and more attention in tracking community.

a) Surfer   b) Dudek

c) Faceocc2   d) Anima

e) Girl   f) Car11

g) Stone   h) Jumping

i) Car4   j) Pktest02
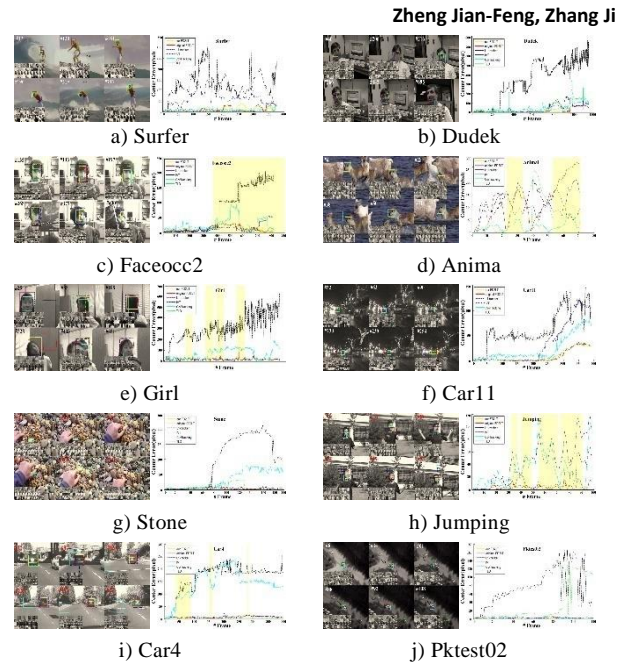
FIGURE 8 Tracking results with six tracking methods. Yellow regions means that, in these frames, trackers lose the targets

Table 1 is the tracking cost comparison of six algorithms used in our experiments. One can see that, our two FD^2LT frameworks work much faster(average 400 times faster under the same experimental settings described in Section 4.1) than l_1-*tracker*, because of the much smaller but well-trained dictionary; the improved FD^2LT (FD^2LT_1 in Table 1) is slightly fast than the original one (FD^2LT_2 in Table 1), because of the simpler optimization in Equation (5) than original one in Equation (2). *TLD*, coded with C++, is the fastest algorithm in our comparative experiments; and our improved and original FD^2LT ranking third and fourth in these algorithms. But our methods work much better than *IVT* (ranking second).

Table 1 The tracking cost comparison of six algorithms used in our experiments.

	Surfer	Dudek	Faceocc2	Animal	Girl	Car11	Stone	Jumping	Car4	Pktest02
IVT	**2.8694**	**3.3211**	**2.7886**	**1.8979**	1.6548	**2.7901**	1.2903	1.2682	0.8479	**1.3503**
CovTrack	1.5707	1.2454	1.1278	1.2534	1.2209	1.7041	**1.8890**	**1.4333**	1.3632	1.1906
TLD	3.0124	2.5078	3.4763	4.1023	3.8792	3.8145	4.6451	3.9561	5.7258	6.7894
l_1-*track*	0.0040	0.0016	0.0020	0.0020.5	0.0023	0.0023	0.0031	0.0029	0.0045	0.0033
FD^2LT_1	2.0886	1.8748	1.7103	1.3596	1.6909	1.7473	1.0794	.0.8927	1.1323	1.1247
FD^2LT_2	2.3469	2.3154	1.8979	1.4620	**1.7392**	2.3624	1.1501	1.1084	**1.6092**	1.1931

5 Conclusion and future works

In this paper, we analysis the reasons for the inefficiency of l_1-*tracker* and the insufficient of original *FDDL* proposed by Mei [4] and Yang [5] firstly. Then, in order to overcome these drawbacks, we present a modified version of *FDDL* and validate the numerical convergence of improved *FDDL*, then introduce the original/modified *FDDL* into video tracking, called FD^2LT.In our framework, three important components (object location, training samples selection, and dictionary learning) are introduced and discussed detail in Section3. Our framework combines generative tracking (i.e., *PF* [1]) with discriminant information (i.e., *FDDL*), and experiments demonstrate the effectiveness and efficiency of our trackers. But there are also some aspects required to be studied in future, including:

1) some conclusions proposed in this paper without strict proof, instead of numerical validation, e.g. theoretic convergence of *FDDL*, why improved FD^2LT has stronger discriminant ability than original one;

2) our FD^2LT framework is much faster than the latter, but it is still far away from real-time tracking(more than 20 *fps* in common video sequence). How to accelerate the tracking efficiency is one the further goals of us and even all computer vision researchers.

Acknowledgements

References

[1] Yilmaz A, Javed O, Shah M 2006 *ACM Computing Surveys (CSUR)* **38**(4) 1-45

[2] Feng C, Wang Q, Wang S, Zhang W, Xu W 2011 *Image And Vision Computing* **29**(11) 787-796

[3] Cands E J, Wakin M B 2008 *Signal Processing Magazine* **25**(2) 21-30

[4] Mei X, Ling H B 2011 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(11) 2259-72

[5] Yang M, Zhang L, Feng X, Zhang D 2011 *IEEE 13th International Conference on Computer Vision (ICVV)* Barcelona 543-550

[6] Aharon M, Elad M, Bruckstein A 2006 *IEEE Transactions on Signal Processing* **54**(1) 4311-22

[7] Benoît L, Mairal J, Bach F, Ponce J 2011 *IEEE Conference on Computer Vision and Pattern Recognition* Colorado Springs USA 2913-20

[8] Tosic I, Frossard P 2011 *IEEE Transactions on Image Processing* **20**(4) 921-34

[9] Rosasco L, Mosci M, Santoro S, Verri A, Villa S 2009 *Technical Report MIT-CSAIL-TR-2009-050* MIT

[10] Zhang J, Wang H Y, Chen F H 2013 *Lecture Notes in Computer Science* **7751** 700-10

[11] Ross D A, Lim J, Lin R S 2008 *International Journal of Computer Vision* **77**(1-3) 125-41

[12] Kalal Z, Mikolajczyk K, Matas J 2012 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(7) 1409-22

[13] Porikli F, Tuzel O, Meer P 2006 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2006)* **1** 728-35

Authors

Jian-Feng Zheng, born in March, 1978, Changzhou, China

Current position, grades: lecturer at the School of Information Science & Engineering, Changzhou University.
University studies: M.S. degree in Master of Computer Applications at Nanjing University of Science and Technology, China in 2011.
Scientific interests: intelligent instruments, computer vision, image processing.
Publications: 3 Patents, 6 Papers.

Ji Zhang, born in November, 1981, Changzhou, China

Current position, grades: lecturer in School of Information Science & Engineering, Changzhou University.
University studies: M.S. degree in Control Science and Engineering at Nanjing University of Science and Technology, China in 2006.
Scientific interest: computer vision, image processing, pattern recognition.
Publications: 10 Papers.

Computer and Information Technologies

Implementation of network management software for HINOC

Jingfei Cui[1], Jinlin Wang[2], Zhen Zhang[3]*

[1]*National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Sciences & University of Chinese Academy of Sciences & Academy of Broadcasting Science, Beijing, 100190, China*

[2]*National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China*

[3]*School of Electronic Engineering and Computer Science, Peking University, Beijing, 100871, China*

Abstract

HINOC (High performance Network Over Coax) is Coaxial cable access technology with independent intellectual property rights in China. In the structure of network management software, network topology discovery and maintenance plays an important role in the system. By analysing the traditional algorithms of topology discovery, this paper introduces an algorithm based on SNMP trap and polling. On the basis of test results, the algorithm which has been implemented in HINOC proved to be reliable, efficient and with low-burden.

Keywords: network management, topology management, HINOC, SNMP

1 Introduction

These HINOC (High performance Network Over Coax) is the transport solution to network access between fibre-optic network and home network. It is acknowledged that making use of cable TV network is a good broadband access technology. On one hand, cable TV network has the characteristics of a large covered range and low cost so that we can use the cable television networks as access network to avoid the negative impact of home construction. On the other hand, coaxial cable has good transmission characteristics, such as large capacity and high quality of transmission [1].

The structure of HINOC is shown in Figure 1. It consists of HB (HINOC Bridge) as bridge device and HM (HINOC Modem) as terminal equipment. HB is connected with optic equipment. It can transmit data as the only interface of optic equipment. Meanwhile it can be used as a control node to monitor HM. HM connects with HB through coaxial cable inside the building, and the other end of HM connects with the terminal devices such as television and computer [1].

This paper introduces topology management in the network management software (NMS) of HINOC. The NMS is based on SNMP (Simple Network Management Protocol) and running on the network manager. The remaining of this paper is organized as follows: in Section 2 we will give an overview of the NMS. Algorithm of Topology Management will be illustrated in Section 3. Section 4 presents the experiment results. Finally, conclusion will be given in Section 5.
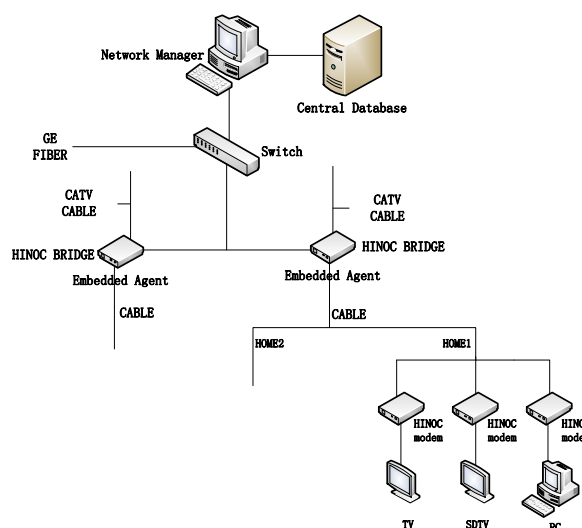
FIGURE 1 HINOC network scenario

2 Network management software of HINOC

The network management software of HINOC is based on the SNMP. The management agents are embedded in the HB and HM. NMS can manage HB and HM with the IP addresses of devices. Every HB is regarded as a unit. It constitutes a subnet with the HM connected with it. The central management software can collect information of HB and HM in a subnet only by the HB instead of communication with the HM directly [2, 3].

Figure 2 is the framework of the central network management based on HINOC. The top-level mainly consists of user interaction interfaces, including management interface and service provided for the WEB. The graphical interface used in the system is developed

Corresponding author e-mail: zhangzhen19910405@163.com

mainly by the Qt Graphical Interface Library. And interface layer responses for analysing the user operation. It concludes the management of topology, performance, fault, configuration and escalation in the NMS (network management system) engine. The NMS engine correspond the operation to the performance function of SNMP, displays the responded data on the GUI and saves them to the database. The FTP is in charge of the system updating management. It is used to storage the latest network management central software and embedded agent of HB/HM.
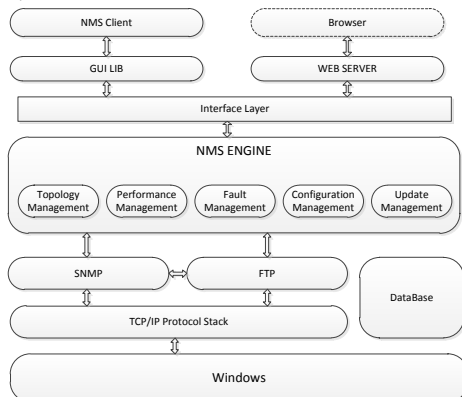
FIGURE 2 Structure of network management software

3 Algorithm of topology management

Text Topology management consists of the process of building topology and maintaining. Considering the importance of topology management, a complete topology management must have 3 aspects: accuracy, commonality and low-burden [5]. We will introduce the traditional algorithm of network topology discovery and the method we used in HINOC.

3.1 TRADITIONAL ALGORITHM OF NETWORK TOPOLOGY DISCOVERY

Current topology discovery methods are shown as follows:
1) ICMP-based method;
2) ARP-based method;
3) Router-protocol (RIP) based method;
4) SNMP-based method [4, 5].

Because of particularity of HINOC network, the traditional algorithm of network topology discovery based on IP network cannot be applied to this system. Firstly, the downlink data is sent by the MAC layer, which has special mechanism of data processing, and we cannot build HB and HM connection by access equipment, for the access equipment does not have routing table. However, the connection built by address forwarding table between MAC addresses cannot meet the demand of management requirement. In this situation, it is impossible to manage the HM by using the IP address. Although polling with broadcasting can build the topology, it can only be used on the same network segment, and will have a huge burden on

the network. Therefore, the algorithm of topology management needs to combine with the characteristics of HINOC MAC layer protocol [6, 7].

3.2 ALGORITHM OF TOPOLOGY MANAGEMENT IN HINOC

The star topology of HINOC is made up of HB and HM. Since the system uses the embedded agent to manage access device by IP address, topology management needs to establish the correspondences between device and IP address. The network management system is implemented by SNMP trap and polling at the network layer. Considering the HM access protocol, this mechanism is designed with 5 notifications of HB's online, HM's online, HM's online reported by HB, HM's offline reported by HB and the shutting down of embedded agent.

Devices that go online have to send SNMP trap with the message composed of node ID, type of device, MAC address, and community. The trap oid of HB's online is 1.3.6.1.3.2.9.1 while HM is 1.3.6.1.3.2.10.3. EMS (Element Management System) receives the trap from HB or HM, acquires the basic information of the device that goes online and saves the source address of the trap. Meanwhile, if the device is HM which has been adopted by a HB, its node ID and MAC address has also been achieved by the HB. HB will send message to EMS to report that the HM belongs to this HB. The trap oid of HM's online reported by HB is 1.3.6.1.3.2.9.4. In this case, EMS will receive two traps which have the same MAC address form HB and HM. As a result, correspondences between HB and HM can be established. Table 1 shows the comparison of different algorithms of topology discovery [4, 5].

TABLE 1 Comparisons of different topology discovery methods

Method	Speed	Accuracy	Burden
ICMP	quick to the alive while slow to others	middle	high
ARP	Quick	low	low
RIP	Middle	middle	low
SNMP	Quick	high	low
Algorithm in HINOC	Quick	high	low

While HB or HM is running, EMS polls periodically to get information of network traffic and output power. If a device does not replay 5 times continuously, EMS could come to a conclusion that the device is offline and modify its state. Besides, information of the device can be saved until it is back online. Figure 3 shows the complete process of topology management.

Even though the traditional algorithm based on SNMP also has the characteristics of accuracy, low-burden and quick-response as we saw in Table 1, it has to send lots of message to every device. With the increasing of devices quantity, the number of messages grows rapidly which can add load on the network [8, 9]. By contrast, the trap sent by embedded agent can avoid this problem and be more efficient. Meanwhile, polling ensures real-time updating of network topology [8-10].
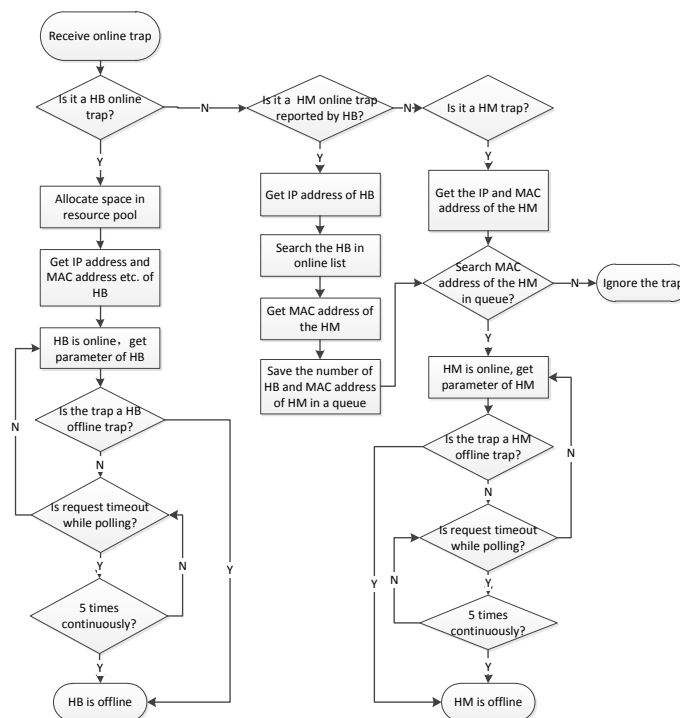
FIGURE 3 shows the complete process of topology management

4 Experiment result

The test environment is shown in Figure 4 and Table 2. Network Management Centre runs on the host computer, and uses the HMD301 Cable Bridge terminal of Shanghai Broadband as Internet access device. The coaxial interface of HB is connected with HM, and the network interface connected with PC. The debugging device connects to HB/HM through the serial port of the host computer, and realizes it with hypertermina.

FIGURE 4 Testing environment

TABLE 2 Parameters of devices

PC	CPU	Intel Core2 DUO E7500
	RAM	2G
	Operating System	Windows XP
HMD301	RAM	32M
	FLASH	32M
	KERNEL	Linux version 2.4.21

4.1 TEST OF ESTABLISHMENT

The Figure 5 shows the process of topology establishment. As shown in the device list on the left, a HM is mounted on the No.1 HB. The right part of the picture shows the basic information of the No.1 HB. From the picture, we can see that both the topological relation and basic information are correct.

FIGURE 5 Main interface of network management software

4.2 MANAGEMENT SCALE TESTING

The Figure 5 shows the testing scale of the Network Management System. Because of the limitation of the testing environment, we cannot test all the 32 HMs actually. So we test with one HB and HM (Table 3). HB and HM sends on-line notification every 15s and the ID of HM increase by one each time to simulate the circumstance that 32 HMs are online. As shown in the Figure 5, when all the 32 HMs are running, the Network Management System is operating smoothly, and we can get the basic information, network flow and output power of the device.

216
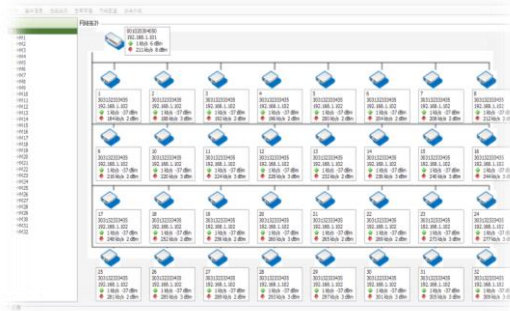
FIGURE 6 Main interface of network management software

TABLE 3 Total time of polling

Parameters	Number of HB/HM	Total time of polling
Network Traffic	1/32	8.6 sec (measured)
and Output Power	225/7200	30 min (estimated)

5 Conclusion and future work

The network management software is developed synchronously with HIONC and it is not complete yet. The network topology based on the MAC layer protocol and HB/HM network structure realizes the management and maintenance of HINOC network. The direct ratio between rotation time and equipment number is still a crucial issue to solve.

Taking the expandability and portability of the network into consideration, the existing C/S architecture cannot meet the requirement that one system could be used anywhere. It could be improved to the B/S architecture in the future and will be able to increase flexibility, shield realization details, provide general interfaces to more manufacturers, and cut down the cost of the system with the help of the Web network management system.

References

[1] Feng O, Jingfei C, Yuping Z, Dou L, Bing Z 2011 Coaxial Cable Access technology in HINOC *Radio and TV Broadcast Engineering* **10** 34-8 *(in Chinese)*

[2] Wenwen C, Xiaohui D, Wu Y, Daoxian W 2012 DESIGN AND IMPLEMENTATION OF HINOC NETWORK MANAGEMENT SYSTEM BASED ON SNMP *Signal Processing, Communication and Computing (ICSPCC 2012) IEEE Press Hong Kong* 527-32

[3] Santosh S C, Madanagopal R 2009 Generic SNMP proxy agent framework for management of heterogeneous network elements *Communication Systems and Networks and Workshops (COMSNETS 2009) IEEE Press Bangalore* 1-6

[4] Yao Z, Jianliang Y, Hua Z 2010 STUDY ON NETWORK TOPOLOGY DISCOVERY IN IPNETWORKS *Broadband Network and Multimedia Technology (IC-BNMT 2010) IEEE Press Beijing* 186-90

[5] Yan H 2012 The study on network topology discovery algorithm based on SNMP protocol and ICMP protocol *Software Engineering and Service Science (ICSESS 2012) IEEE Press Beijing* 665-8

[6] Bing Z, Changming L, Kun W 2012 Theoretical Analysis of the Performance of HINOC MAC Protocol *Journal of Network New Media* **1**(3) 24-9 *(in Chinese)*

[7] Xiaona Z, Minchao W, Qiongqiong W, Bing Z 2013 Design and Validation of HINOC MAC Signaling Protocol *Journal of Network New Media* **2**(2) 18-21 *(in Chinese)*

[8] Huanran W, Mingwei X 2004 Survey on SNMP Network Management *Mini-Micro Systems* **25**(3) 358-60 *(in Chinese)*

[9] Chunling C, Guoliang C, Zongjian S 2010 SNMP-based broadcasting parallel polling algorithm for network management *Application Research of Computers* **27**(12) 4711-12 *(in Chinese)*

[10] Douglas M, Kevin S 2005 *Essential SNMP (2nd Edition)* O'Reilly: America 141-53

Authors

Jingfei Cui, born on April 18, 1973, Shandong, China

Current position, grades: Vice Chair of Focus Group Smart Cable TV, ITU-T SG9, Director of Cable Network Institute, SARFT.
University studies: Ph.D. student of National Network New Media Engineering Research Center, Institute of Acoustics.
Scientific interest: broadband network, information security, interactive service infrastructure.

Jinlin Wang, born on December 13, 1964, Tianjin, China

Current position, grades: professor, doctor supervisor, director of Network and New Media Technology Research Center.
University studies: Master degree, Institute of Acoustics, Chinese Academy of Sciences, 1986-1989.
Scientific interest: digital signal processing, application of DSP, digital TV source and channel decoding technology and receiving system.

Zhen Zhang, born on April 5, 1991, Anhui, China

University studies: Master student of Peking University.

Evaluating method of DB contracting based cloud model and Gray relational degree

Zhiding Chen, Qilun Zou[1*], Qian Liu

College of Hydraulicand Environmental Engineering, China Three Gorges University, Yichang 443002, China

Abstract

The proposed evaluating method based Cloud Model and Gray Relational Degree aims to solve the fuzziness and randomness problems of evaluation methods and the outcome efficiently. Cloud model is implemented to convert qualitative concept into quantitative value based index system. Gray Relation Degree theory is implemented to access to evaluation index weights.

Keywords: Cloud Model, Gray relational degree, DB contracting, evaluation of bid, weight

1 Introduction

With economic development, technology advances, project scale expansion, developer' need for overall management, DB contracting model is increasingly favoured by the market. It is put forward new challenges for Project Bidding. Bidding of DB model exist fuzziness and randomness problems which are due to inaccurate information such as status of market supply demand, value orientation of owner, characteristics of the project subject to tenderers, and bidder's strategy. Further study of bidding institutional innovation and theoretical methods are needed.

Bid evaluation is the most important part of tendering activities. How to select bid evaluation indicators and their weights is a difficulty. Many researchers have done lots of studies and promoted corresponding theory methods and assessment models, including: Value Engineering [1], analytic hierarchy process [2], comprehensive fuzzy evaluation method [3], grey correlation analysis [4], principal component analysis [5], entropy method [6], ANN [7], DS evidence theory [8], set pair theory [9] and other single evaluation method, and combination of above methods, such as: AHP and fuzzy comprehensive evaluation [10], DEA and information entropy method [11] etc.

The above mentioned evaluation methods are suitable for those project which have the characteristics of adequate preparation, concentrated information, low risk, single target. However, cannot reflect the randomness and fuzziness and their relation of bid evaluation of DB project. Cloud model can realize the conversion between the qualitative and quantitative description that can use single rule uncertainty reasoning to quantify evaluation indicators; grey correlation method has the characteristics of high utilization of information, weight calculation reasonable and can obtain the weight of each evaluation indicators. We combine the advantages of cloud model and grey correlation method to carry out the study of project evaluation method.

2 Cloud model descriptions of bid evaluation indicators of DB project

2.1 CONCEPT OF CLOUD MODEL

2.1.1 Definition of Cloud

Let X denote an ordinarily set, $X \in \{x\}$, which is called a domain. $\tilde{A}$ is the fuzzy subset on the domain X, which means there always exists a stable numerical variable $u_{\tilde{A}}(x)$, which is called the element's x membership degree on $\tilde{A}$. If the elements from domain are simple and orderly, X is underlying variable. The distribution of membership degree on X is called Membership Cloud. If the elements are not simple and orderly, x, according to a rule, can be mapped to another orderly universe, the one and only one and corresponds, for the basic variables, membership in the distribution called membership cloud [14].

2.1.2 Numerical characteristic of Cloud

1) Membership cloud employs expectation, entropy and hyper-entropy to describe a specific concept.
2) Expectation (Ex) expresses the point which is the most suitable to represent the domain of the concept and it is the most typical sample after this concept to quantify.
3) Entropy (En) reflects qualitative concept. The uncertainty is reflected from three aspects:

    a) Entropy reflects the range of domain space, which could be accepted by the specific concept.

    b) It reflects the probability of cloud droplet represents linguistic terms in domain space.

    c) It can be used to express the relationship between

randomness and fuzziness. Entropy represents a granularity of a concept, which could be measured.

4) Hyper entropy (He) describes the uncertain measurement of entropy. It reflect the coherence of uncertainty.

2.1.3 Cloud generator

Cloud generator means generating algorithm of cloud. Cloud generator includes forward cloud generator and backward cloud generator. Forward cloud generator can be divided into X condition cloud generator and Y conditions cloud generator. Forward cloud generator produces cloud droplets described by the specific concept. X condition cloud generator employs numerical characteristics Expectation, Entropy, Hyper entropy and specific value x_0 to generate series droplets. Y condition cloud generator employs numerical characteristics Expectation, Entropy, Hyper entropy and specific value u_0. These two conditions cloud generators are the basement of uncertain

reasoning which can achieve conversion between qualitative and quantitative.

2.2 SINGLE CLOUD MODEL UNCERTAINTY REASONING RULES

A formal description of qualitative rules: If A then B. A and B are cloud objects described in language value. Cloud generators are the basement of uncertain reasoning by using cloud model. A single rule generator consist of X condition cloud generator and Y condition cloud generator. CGA represents the X condition cloud of input language value A, CGB represents the Y condition cloud of output language value B. When a given value is input to stimulate CGA, CGA randomly generates a set of values u_i (which reflects it is intensity of activation towards qualitative rules). By inputting u_i in CGB, CGB generates a random cloud droplets $C_{drop}.(y_i, u_i)$. It should be noted that in order to achieve uncertain reasoning the cloud droplets and output values are not unique and determined [15]. The model of Single Rule Generator is shown in Figure 1.
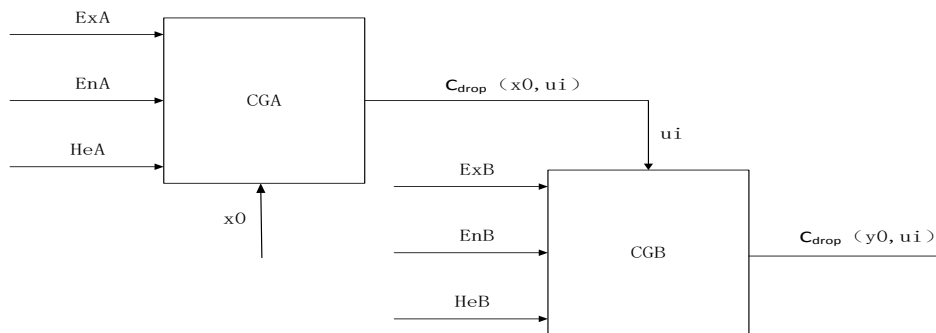
FIGURE 1 Single Rule Generator

2.3 ESTABLISHMENT OF BID EVALUATION INDEXES BASED CLOUD MODEL

Bid evaluation indicators of DB projects should have the characteristic of testability, completeness, independence,

sensitivity and consistency. Reasonable selection of indicator system will directly affect credibility of the final results. Following indicators system in Table 1 is summarized based on the study of evaluation methods of 169 bidding documents.

TABLE 1 Bid evaluation indicators system

Target	First-grade indicators	Second-grade indicators
Bid evaluation indicators system	Bidder	Economic strength
		credit
	Project management team	Performance and quality of project manager
		Quality and experience of project management team
	Design	Designing idea and method
		Structure, layout and function of drawings
		Specification and requirement of design
	Construction design	Construction design and technical measure
		Schedule, quality plan and assurance measures
	Price	Offer
		Correctness and completeness of bidder
		Rationality of offer

As it is seen from Table 1, bid evaluation indicators of each second-grade indicators are divided into several comments according to the experience of experts. Takes Economic strength of tenderer for example, economic strength will be classified as "strong, less strong, average,

less weak, weak" according to the enterprise scale, amount of fixed assets and guarantee funds. Cloud model is expressed as: "(2100, 300 / 3, 10), (1600, 250 / 3, 10), (1200, 200 / 3, 10), (800,250 / 3, 10), (300,300 / 3, 10)".
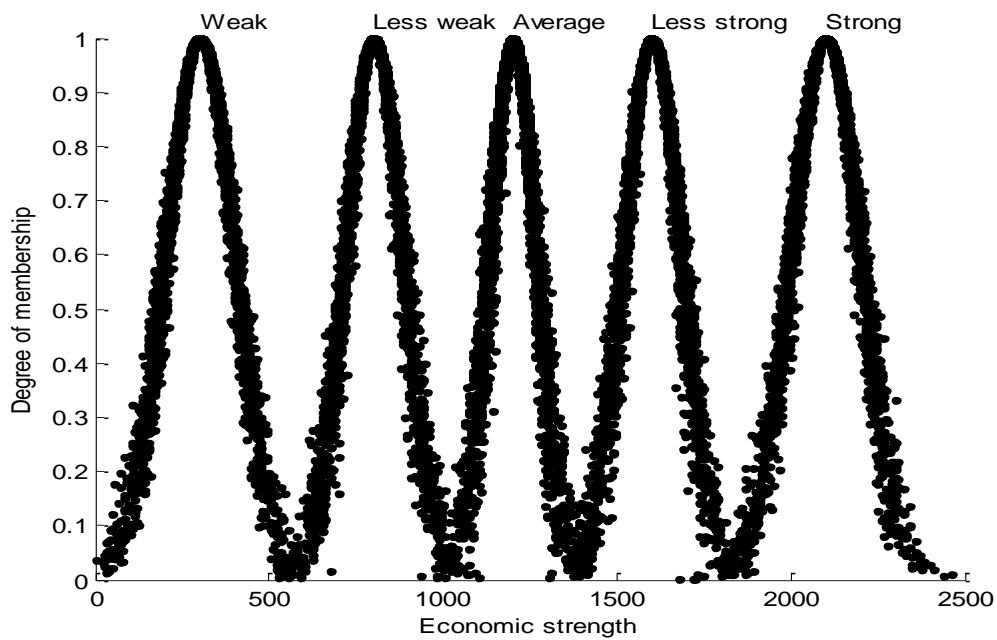
FIGURE 2 Cloud droplets of Economic strength and its membership degree

3 Experimentation

3.1 EXAMPLES OF BASIC DATA

Enterprises A, B, C, D participate in the bidding of a DB

project.evaluation experts grade each enterprises on bid evaluation indicators system. As showed in Figure 2, high value of one indicator shows that the corresponding program is better than other programs.

TABLE 2.Indicator evaluation result of tenderer

Second-grade indicators	A	B	C	D
Economic strength	920	1567	1103	2050
Credit	Very good	Good	Good	Average
Performance and quality of project manager	Average	Very good	Very good	Good
Quality and experience of project management team	Not good	Very good	Average	Good
Designing idea and method	8	5	6	9
Structure, layout and function of drawings	7	4	9	8
Specification and requirement of design	Satisfied	Average	Less satisfied	Average
Construction design and technical measure	6	8	5	4
Schedule, quality plan and assurance measures	8	5	7	6
Offer	3194	3704	3657	3235
Correctness and completeness of bidder	Complete	Incomplete	Average	Complete
Rationality of offer	Reasonable	Unreasonable	More reasonable	Average

3.2 ESTABLISHMENT OF CLOUD MODEL AND UNCERTAINTY REASONING

Step1: Establish second degree indicators comment sets according to the nature of the evaluation indicators. Comment sets are divided into five levels and can be used to describe the inherent ambiguity of indicators. As seen from 1.3 establishment of bid evaluation indications based cloud model.
Step2: Develop grade comment set and grade cloud model. We adopt centesimal system to measure the level of indicators. Qualitative comments "Very Good, Good, average, Not Good, bad" is described as "(90,10/3,0.3), (75,10/3,0.3), (60,10/3,0.3), (45,10/3,0.3), (30,10/3,0.3) in cloud model."

Step3: By using the single rule uncertainty reasoning of cloud rules to transform qualitative description to quantitative description. Level of qualitative indicators is quantized to a specific value. We build up following rules: A case study of economic strength of tenderers.

If economic strength is "strong" then score is "high". If economic strength is "less strong" Then score is "less high". If economic strength is "average" Then score is "medium". If economic strength is "less weak" then score "low". If economic strength is "weak" then score "lower".

Using the single rule uncertainty reasoning of cloud model and MATLAB software, we can obtain indicator grades as shown in Table 3:

TABLE 3 Indicators grades of tenderers

Second-grade indicators	A	B	C	D
Economic strength	49.57	73.65	54.31	88.61
Credit	89.74	77.52	76.39	62.03
Performance and quality of project manager	63.40	74.55	91.61	78.14
Quality and experience of project management team	43.29	78.96	58.34	89.05
Designing idea and method	88.71	56.30	67.47	94.18
Structure, layout and function of drawings	74.82	42.37	90.35	84.76
Specification and requirement of design	91.25	62.49	74.11	59.89
Construction design and technical measure	67.81	84.59	50.73	42.30
Schedule, quality plan and assurance measures	85.16	57.77	74.38	69.02
Offer	90.53	66.19	70.34	87.46
Correctness and completeness of bidder	88.33	47.58	61.04	89.07
Rationality of offer	73.67	43.50	91.18	60.08

3.3 CALCULATE INDEX WEIGHTS BY GRAY RELATIONAL DEGREE

Delphi method, AHP and entropy weight method are the main methods to determine index weights. In this paper, we adopt grey related degree method to determine the index weight by considering the randomness and fuzziness of evaluation indicators. First, initialize indicators values in Table 3 to get a new Data Matrix, then calculate the grey relational grades of P_i and other indicators (except P_i) [16],

$$r(P_i, \overline{P_i}) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} r(P_i, P_j), \qquad (1)$$

$$r(P_i, P_j) = \frac{1}{m} \sum_{h=1}^{m} r(x_i(h), x_j(h)), i, j = 1, ..., n, \qquad (2)$$

$$r(x_i(h), x_j(h)) =$$

$$\frac{\min_{i,j \in n} \min_{h \in m} |x_i(h) - x_j(h)| + \xi \max_{i,j \in n} \max_{h \in m} |x_i(h) - x_j(h)|}{|x_i(h) - x_j(h)| + \xi \max_{i,j \in n} \max_{h \in m} |x_i(h) - x_j(h)|}, \qquad (3)$$

$$w_i = \frac{r(P_i, \overline{P_i})}{\sum_{i=1}^{n} r(P_i, \overline{P_i})}. \qquad (4)$$

Finally, normalizes this n groups grey relational degree and we can get the relative weight of each indicators, namely. According to the above calculation, we can obtain indicators weight at various levels. As shown in Table 4:

TABLE 4 levels of index weights

First-grade indicators	Weight	Second-grade indicators	Weight
Bidder	0.1610	Economic strength	0.0729
		Credit	0.0880
Project management team	0.1408	Performance and quality of project manager	0.0702
		Quality and experience of project management team	0.0706
Design	0.2635	Designing idea and method	0.0881
		Structure, layout and function of drawings	0.0856
		Specification and requirement of design	0.0897
Construction design	0.1742	Construction design and technical measure	0.0850
		Schedule, quality plan and assurance measures	0.0892
Offer	0.2605	Price	0.0871
		Correctness and completeness of bidder	0.0869
		Rationality of price	0.0865

4 Results and conclusions

From Table 4 we can see that weight of design programming is the highest which indicating designs programming is the most important evaluation indicator during bid evaluation of DB project. The next highest weight is offer, which is the core-competitiveness indicator based on design programming. Weight of offer is slightly less than design programming's. The sum of this two weights is 0.524, which is the decisive factor in winning the project; Bidder and project management team cannot be ignored, which is the organizational measures to achieve overall objective of project.

Case study results show that the Cloud model is practicable for the conversion between the qualitative and quantitative description of DB bid evaluation indicators.

Combined method of single cloud model uncertainty reasoning rules and grey correlation degree can quantitatively calculate the weight of bid indicators. This method fully reflects the randomness and fuzziness evaluation of bid evaluation and overcomes the subjective and arbitrary determination of indicators weights. The evaluation model proposed in this paper is practical and deserving extensive use.

References

[1] Xiang W 2004 Value engineering applied in the bidding *Journal of Chongqing University, Natural Science Edition* **6** 148-150 *(in Chinese)*

[2] Shen L, Li Q 2005 Application of Analytic Hierarchy Process (AHP) in Construction Engineering Bid Evaluation *Construction Technology* **2** 64-6

[3] Zang X, Li P, Zhang J, Ding S 2007 Application of Improved Fuzzy Integrating Assessment in Evaluation *Journal of Jiangsu University of Science and Technology Natural Science Edition* **21**(6) 9-12 *(in Chinese)*

[4] Wang S, Zhang Y, Cao Y, Hao C 2005 Application of multiple-attribute gray incidence decision-making model based on a linear transformation operator on [-1, 1] to bid appraisal for the project *Donghai Marine Science* **23**(4) 49-53 *(in Chinese)*

[5] Chen T 2005 Model of Tender Evaluation Based on Principal Component Analysis *Journal of Wuhan University Natural Science Edition* **51**(S2) 54-6 *(in Chinese)*

[6] Lu S, Huang Q, Sun X 2010 The applied research of entropy proportion in the tendering procedures Journal of Hydroelectric Engineering, 2010 **29**(3) 221-4 *(in Chinese)*

[7] Xiao K, Xia T 2006 Application of neural network to comprehensive evaluation of urban financial development *Engineering Journal of Wuhan University* **39**(6) 122-4 *(in Chinese)*

[8] Chen G, Wang J, Li R, Hu S 2012 Evaluation of hydraulic project tender by D-S evidence theory *Journal of Hydroelectric Engineering* **31**(3) 263-6 *(in Chinese)*

[9] He M, Liu L, Wang H, Liu J 2012 Unknown weight multiple attribute decision for engineering appraisal bidding based on set pair analysis *Journal of Central South University, Science and Technology* **43**(10) 4057-62 *(in Chinese)*

[10] Zhang M, Liu Z 2008 Application of AHP and Fuzzy Comprehensive Assessment in Tender Assessment of Scientific and Technical Projects *Fire Control and Command Control* **33**(10) 124-6 *(in Chinese)*

[11] Cao L, Liu B, Wang X, Feng T 2011 A Study on the Bidding Evaluation Method Based on DEA and Improvement of the Information Entropy *Journal of Chongqing University, Social Science Edition* **17**(2) 86-9 *(in Chinese)*

[12] Li D, Liu C 2004 Study on the universality of the normal cloud model *Engineering Science* **6**(8) 28-34 *(in Chinese)*

[13] Ren H, Yan Y, Zhou T, Xiang X, Zhang Y 2011 Evaluation on cooperative partners in organization coalition for mega projects based on cloud model and gray correlation analysis *China Civil Engineering Journal* **44**(8) 147-52 *(in Chinese)*

[14] Li D, Meng H, Shi X 1995 Membership clouds and membership cloud generators *Journal Of Computer Research And Development* **32**(6) 16-8 *(in Chinese)*

[15] Hu S, Li D, Liu Y, Li D 2006 Mining Weights of Land Evaluation Factors Based on Cloud Model and Correlation Analysis *Geomatics and Information Science of Wuhan University* **31**(5) 423-6 *(in Chinese)*

[16] Luo B, Yuan K, Sui L, Ma X 2002 DGR-based investment decision model with application *Systems Engineering-Theory Practice* **22**(9) 132-136 *(in Chinese)*

Acknowledgments

Authors

Zhiding Chen, born in May, 1974, Hubei, China

Current position, grades: Doctor of engineering, associate professor and master's supervisor in China Three Gorges University.
University studies: PhD in Construction and Management of Hydraulic Engineering in Wuhan University.
Scientific interests: building information modeling, construction management and cost engineering.
Publications: 1 Monograph, 30 Papers.

Qilun Zou, born in August, 1988, Hubei, China

Current position, grades: Master of Engineering Management in China Three Gorges University.
University studies: the bidding and project management in China Three Gorges University.
Scientific interests: risk evaluation of EPC general contracting mode.

Qian Liu, born in December, 1987, Hubei, China

Current position, grades: Master of Engineering Management, lecturer in China Three Gorges University.
University studies: Management Science and Engineering in China Three Gorges University.
Scientific interest: international project management.
Publications: 4 papers.

Computer and Information Technologies

WSN image acquisition method based on interleaving extraction and block compressed sensing

Qiang Fan[1, 2], Dongjian He[1*], Min Zhang[3]

[1]*College of Mechanical and Electronic engineering, Northwest A&F University, Yangling 712100, Shaanxi, China*

[2]*College of Water Resources and Architectural, Northwest A&F University, Yangling 712100, Shaanxi, China*

[3]*College of Science, Northwest A&F University, Yangling 712100, Shaanxi, China*

Abstract

Aiming at disadvantages of current wireless sensor network in the aspect of image acquisition, this thesis proposes a WSN image acquisition method based on Interleaving Extraction and Block Compressed Sensing (IE-BCS). The method uses interleaving extraction to divide an original image into several sub-images at an encoding terminal, then compressive sampling and encoding for each sub-image by means of observation matrix weighted BCS and transmits data to a decoding terminal by their own independent channels. Next, the decoding terminal chooses corresponding decoders according to reception situations and reconstructs the original images by solving sparse optimization problems. Experimental results show that the method can save hardware resources effectively and improve robustness of image transmission.

Keywords: compressed sensing, interleaving extraction, observation matrix, block strategy

1 Introduction

As Wireless Sensor Network (WSN) develops and quantity of information requirements grows increasingly, how to transmit images and video signals with high quality in network seems to be quite important. The current image acquisition method based on WSN utilizes a digital camera to acquire image information, compresses images by JPEG encoding algorithm and transmits data by wireless communication after segmenting and packing them. Then, data arrive at a user terminal after being transmitted by many nodes. Next, the user analyses the received data package in order to recover original images. Although the foregoing method can realize image acquisition and wireless transmission favourably, the following problems still exist:

1) image sampling follows Nyquist sampling theorem, which results in the situation that the size of data stored and processed by sensor nodes is large and increases nodes' consumption of power supply;

2) if there are several image sampling nodes in the network, data transmission quantity in the network will be increased sharply, which will easily lead to problems, for instance, network congestion; and

3) data of wireless transmission tend to be affected by environment and energy factory easily, so users cannot recover original images correctly if situations like error codes or packet loss appear in transmitting procedure.

Directing at deficiencies of WSN in the aspect of image acquisition, this thesis proposes a WSN image acquisition method based on Interleaving Extraction and Block Compressed Sensing (IE-BCS). Firstly, the method carries out interleaving extraction for an original image to obtain several mutually independent descriptions. Then, it uses the BCS method of the weighted observation matrix to implement compressive sampling and encoding for each description, and transmits them to a decoding terminal by their own independent channels. Next, the decoding terminal chooses corresponding decoders according to reception situations and reconstructs the original image by solving sparse optimization problems. Because the method utilizes compressed sensing (CS) technology, it can effectively reduce data sizes of network transmission, decrease pressure on nodes' computation and storage and lower nodes' consumption of power supply. At the same time, each description can independently recover images with acceptable quality, which not only improves robustness of image transmission but also solves the problem that images cannot be reconstructed correctly because of packet loss in the transmitting procedure.

2 Compressed sensing theory

Compressed Sensing theory is a new signal acquisition theory proposed by D. Donoho and E. Candès et al. in 2006 [1-3]. The theory indicates that when the signal is sparse or compressible in a transform domain, the measurement matrix non-coherent of transformation matrix can be used to linear project transform coefficient into the low dimensional observation vector. This projection maintains the required reconstruction signal information, and through further solving sparse optimization, can precisely

or high probability accurately reconstructs the original high dimensional signal from the low dimensional vector. The CS theory properly compresses data when signals are acquired so that original signals can be reconstructed accurately by a few sample points, which will largely reduce sampling frequency of low image signal and cost of data storage and transmission.

The specific process of compressed sampling is shown in Figure 1. It is assumed that there is a signal $X \in R^n$. If the signal X is compressible in a certain orthogonal basis or a compact framework Ψ, get the transformation coefficient $\Theta = \Psi^T X$. Use a measurement matrix $\Phi \in R^{m \times n}$ ($m << n$), which satisfies restricted isometry property (RIP), to carry out linear transformation for the coefficient vector Θ and obtain the transformed vector $Y \in R^m$. Then, an optimization solution method can be used to reconstruct the original signal X from the vector Y accurately or high probably.
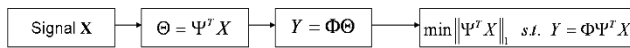
FIGURE 1 The specific process of compressed sampling

3 Key technology of the method based on IE-BCS

3.1 INTERLEAVING EXTRACTION

The principle of interleaving extraction is that adjacent pixels in an image are equally distributed to different descriptions to the largest extent. The interleaving extraction method shown in Figure 2 is used to segment an original image into 2 or 4 sub-images and each sub-image is corresponding to one description.

a) 2 sub-images   b) 4 sub-images
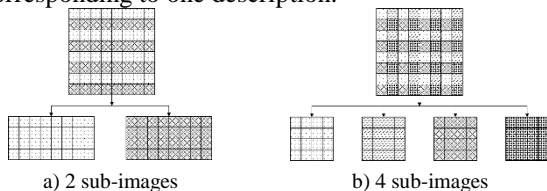
FIGURE 2 Image division process by interleaving extraction

3.2 THE BCS OF WEIGHTED OBSERVATION MATRIX

Application of compressed sensing theory in image signal is commonly used in block compressed sensing method [4, 5]. The conventional BCS divides a whole image into some equal-sized sub-blocks, and then observes and constructs each sub-block independently. It owns the advantage of small observation matrix and of low computation complexity. To specify, the conventional BCS first divides an image of $I_m \times I_n$ into sub-blocks and assumes the size of each sub-block as $n_B \times n_B$, then scans each sub-block and generates the vector X_i containing N_B ($n_B \times n_B$) elements, and next observes them with an identical observation matrix $\Phi_B \in R^{M_B \times N_B}$ ($M_B < N_B$)

and gets the observation value $Y_i \in R^{M_B}$, that is, $Y_i = \Phi_B X_i$. After the observation value gained, all of the sub-blocks are restored with the construction algorithm, and last are combined to form up the whole image.

As Gaussian random matrix is irrelevant with most matrixes of fixed orthogonal basis [3], the conventional BCS means usually applies independent and identically distributed Gaussian random matrix as the observation matrix. But the measurement disregards the relationship of the reconstruction precision of the images with different frequencies to the overall quality of the reconstructed images, and thus results in reconstructions of poor quality. The BCS of weighted observation matrix proposed in this research is an improvement to the conventional BCS, and is a way of weighted processing the elements of different frequencies in the observation matrix to make the images of low frequencies higher reconstruction precision, and to improve the overall quality of the reconstructed image.

3.2.1 Weighted Processing of the observation matrix

According to the characteristics of human visual system, human eyes have greater sensitivity to the low frequency components of the image data. Therefore, the reconstruction error in low frequency components determines the quality of the reconstructed image. In accordance with the principle of the DCT transform, The DCT low frequency coefficient of an image distributes in the upper left corner of the DCT coefficient matrix, while the high frequency coefficients in the lower right corner, and the main energy of the image is concentrated in the low frequency regions of the image. Therefore, the JPEG quantization table got processed, generating weighted coefficient matrix, which was then Zig-Zag ordering, creating weighted coefficient curve, as shown in Figure 3. The weighted coefficient curves reveals: when weighted processing each element in the observation matrix, the larger the weighted value of the matrix elements computing the data in the low frequency positions of the image. The higher the reconstruction precision of the image data in the low and medium frequency part; but the weighted value of the matrix elements computing the data in the high frequency positions of the image is comparatively small, and the construction errors are comparatively bigger. As the low frequency components are the main factors affecting the image quality, this method can make the overall quality of the reconstructed image improved.

a) 8×8 blocks   b) 16×16 blocks
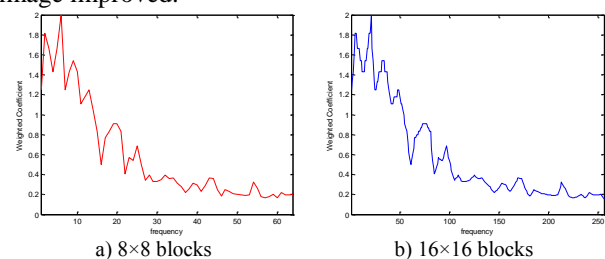
FIGURE 3 Weighted coefficient curve of different blocks

When weighting the observation matrix, first the weighted coefficient should be sequenced by Zig-Zag in the same way as image sub-blocks generate vectors, forming the weighted coefficient vector W_B in the length of N_B and the vector element $w_B(n)$. Then each element in each line of the original measurement matrix Φ_B multiplies each element in the corresponding position of the weighted vectors:

$$\phi_B'(m,n) = \phi_B(m,n)w_B(n), (m=1,2,...,M_B, n=1,2,...,N_B) \quad (1)$$

and thus, forming the weighted measurement matrix Φ_B'. As to a whole image, the interrelationship among the observation matrix, the image data and the observation values is as Equations (2), among which $l = (I_m \times I_n)/N_B$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_l \end{bmatrix} = \begin{bmatrix} \Phi_B' & 0 & ... & 0 \\ 0 & \Phi_B' & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \Phi_B' \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_l \end{bmatrix}. \quad (2)$$

3.2.2 Contrast of the reconstruction results before and after the observation matrix weighting.

Lena, Cameraman, Boats and Fruits grey images in the size of 256×256 selected as the test images, the reconstruction results by Gauss random matrix Φ_B and weighted measurement matrix Φ_B' were analysed. The experiment adopted DCT to transform the sparsity of the images, and l_1 − minimization algorithm to reconstruct the images, and the simulation were divided into two block sizes of 8×8 and 16×16. The simulation results are as Figure 4.

a) Lena b) Cameraman

c) Boats d) Fruits
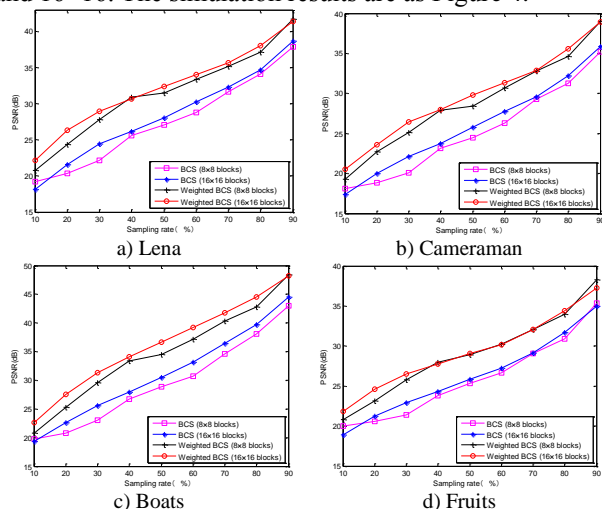
FIGURE 4 The results of reconstructed images by different observation means

From the comparison between the reconstruction results by BCS before and after the weighted observation matrix at an identical sampling rate, such can be observed that the PSNR of the reconstructed grey images of Lena,

Cameraman, Boats and Fruits through weighted observation matrix have been raised 4dB, 3.7dB, 5.1dB and 3.1dB respectively than the average, and among them the reconstruction quality of the 16×16 block is better than the 8×8 block. This method proposed simply processes the measurement matrix in a weighted way but does not add to the computation complexity, and proves to enable the reconstruction precision to greatly increase with the sampling rate unchanged.

4 IE-BCS system frame work

IE-BCS system is composed of an encoding module and a decoding module. 4-descriptions are used as an example to introduce composition and specific functions of each module.

4.1 ENCODING MODULE

The function of the encoding module is that it implements interleaving extraction and BCS encoding for the original image, as shown in Figure 5. The specific processing procedure is shown as follows.
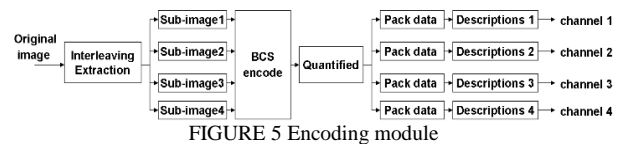
FIGURE 5 Encoding module

1) Implement interleaving extraction for the original image to generate 4 sub-images.

2) Divide each sub-image into sub-blocks with the size of 8×8 or 16×16, carry sparse conversion for each sub-block and use a weighted measurement matrix to sample and encode the transformation coefficient. When an encoder/ decoder is designed, seeds generating random matrix are fixed at both ends to ensure that the same measurement matrix is used when observation is carried out at the encoding terminal and reconstruction is performed at the decoding terminal.

3) Pack data after being quantified and transmit them into the decoding terminal by their own independent channels. It is worth mentioning that a uniform quantification method is applied to the process of quantification.

4.2 DECODING MODULE

The function of the decoding module is that it receives and analyses data packages in channels, and select corresponding decoders according to reception situations to decode each description, as shown in Figure 6. The specific processing procedure is shown as follows.
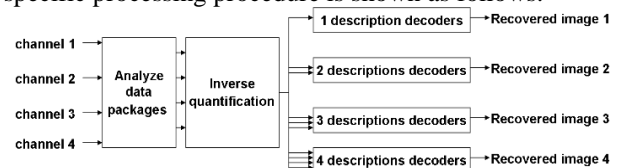
FIGURE 6 Decoding module.

Fan Qiang, He Dongjian, Zhang Min

1) Receive data packages in channels and select corresponding decoders according to reception situations. The principle by which decoders are selected is that description decoders will be selected if complete descriptions are received. In another word, if each description is received completely, decode each description, respectively, combine them crosswise and obtain complete images. Otherwise, decode complete descriptions only and use interpolation algorithm to recover approximate images.

2) Analyse data packages and process them by inverse quantification to obtain an observed value of the transformation coefficient.

3) Reconstruct the original transformation coefficient from the observed value by solving sparse optimization, and then use inverse sparse conversion to reconstruct the original image.

5 Simulation experiment

In MATLAB environment, Lena, Boats, Barbara and Peppers grey images in the size of 512×512 are selected as the test images. Utilize the interleaving extraction to divide the original image into 4 sub-images, segment each sub-image into sub-blocks with the size of 16×16, use DCT to carry out sparse conversion for sub-blocks, adopt the weighted measurement matrix as a measurement matrix and use l_1 – minimization algorithm for reconstruction the images. The reconstruction results of the method based on IE-BCS at different sampling rates are shown in Figure 7. Results of 4-description reconstruction images with different sizes are presented in Figure 8.
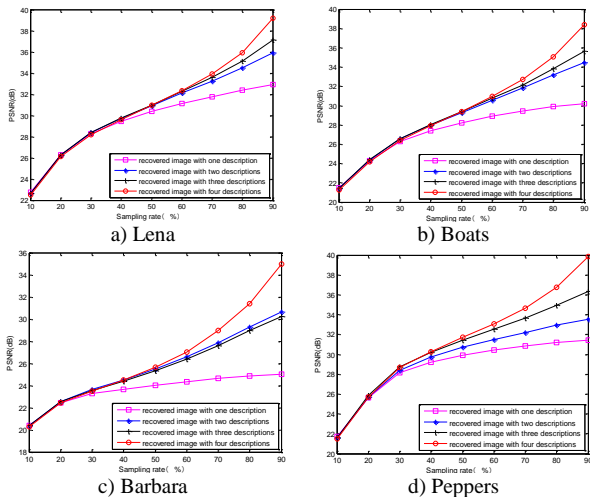
FIGURE 7 Reconstruction results of the method based on IE-BCS at different sampling rates

a) Lena b) Cameraman
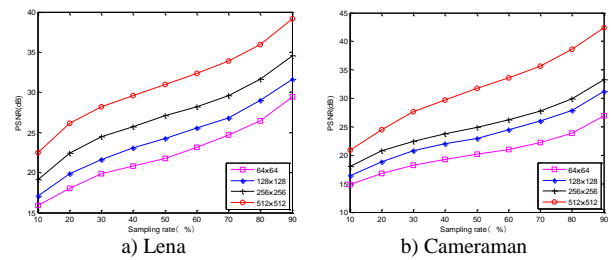
FIGURE 8 Results of 4-description reconstruction images with different sizes

Test results show that reconstruction quality (PSNR) of images have close relationship with the sampling rate and the number of descriptions, which are received completely. When the sampling rate is lower than 30%, the reconstruction quality of images obtained by each decoder is basically the same. As the sampling rate is improved, the reconstruction quality of images is improved obviously. When the sampling rate is higher than 30%, the larger the number of descriptions, which are received completely, is, the better the reconstruction quality of images is. When the sampling rate exceeds 70%, the original image can be reconstructed favourably on the premise that only one complete description is needed. According to Figure 8, it can be known that the larger the image size is, the higher the smoothing degree of sub-images generated by interleaving abstraction is, and the better the reconstruction effect is. Since the method based on IE-BCS adopts a block processing method and ensures complexity of the observation process will not be changed with the image size, the method is more appropriate for processing high-definition images.

6 Discussion and conclusion

This thesis proposes a WSN image acquisition method based on IE-BCS. The methods utilizes the block compressed sensing theory and the interleaving extraction technology, so it can reduce pressure on nodes' computation and storage effectively and decrease data sizes of network transmission. In addition, this thesis also improves the traditional BCS method and puts forward the BCS method weighted by a measurement matrix, which enhances precision of image reconstruction. Meanwhile, it uses 4 international standard test images to carry out an experiment. Results indicate that the method can save hardware resources efficiently and improve robustness of image transmission.

Acknowledgments

References

[1] Candès E 2006 Compressive sampling *Proceedings of International Congress of Mathematics* Madrid Spain **3** 1433-52
[2] Donoho D L 2006 Compressed sensing *IEEE Transactions on Information Theory* **52**(4) 1289-306

[3] Shi G M, Liu D H, Gao D, Liu Z, Wang L J 2009 Advances in theory and application of compressed sensing *Acta Electronica Sinica* **37**(5) 1070-1081 *(in Chinese)*

[4] Lu G 2007 Block compressed sensing of natural images *Proceedings of the 15th International Conference on Digital Signal Processing* Cardiff UK 403-6

[5] Li Y 2011 Improved model of image block compressed sensing. *Computer Engineering and Applications* **47**(25) 186-9

[6] Baraniuk R G 2007 Compressive sensing *IEEE Signal Processing Magazine* **24**(4) 118-121

[7] Candès E, Romber J, Tao T 2006 Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information *IEEE Transactions on Information Theory* **52**(2) 489-509

[8] Li S T, Wei D 2009 A survey on compressive sensing *Acta Automatica Sinica* **35**(11) 1369-77 *(in Chinese)*

[9] Liu D J, Shi G M, Zhou J S et al. 2009 New method of multiple description coding for image based on compressed sensing *Journal of Infrared and Millimeter Waves* **28**(4) 298-302 *(in Chinese)*

[10] Cheng Y 2012 Adaptively encoding of compressive sensing for image signal Hubei: *Master thesis of South-Central University for Nationalities(in Chinese)*

[11] Zhao C H, Liu W 2011 Image Multiple Description Coding Method Based on Interleaving Extraction and Block Compressive Sensing Strategy *Journal of Electronics & Information Technology* **33**(2) 461-5

Authors

Qiang Fan, born in September, 1980, China

Current position, grades: doctor candidate in College of Mechanical and Electronic Engineering, Northwest A&F University.
Scientific interests: wireless sensor network and compressed sensing.

Dongjian He, born in August, 1957, China

Current position, grades: doctor of engineering, professor, doctor candidate instructor, dean of College of Mechanical and Electronic Engineering, Northwest A&F University.
Scientific interests: image analysis and recognition, intellectualized detection and monitoring, multimedia network applications.

Min Zhang, born in February, 1982, China

Current position, grades: doctor candidate in College of Science, Northwest A&F University.
Scientific interests: environmental biophysics and model.

Research on cloud storage technologies of typical crop growth environment monitoring data

Yuansheng Wang, Huarui Wu*, Lei Luo, Weihong Ma, Kun Liu

[1]*Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China*

[2]*Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China*

Received 2 August 2014, www.cmnt.lv

Abstract

The typical crop is an important part of regional economy, using of information technology tools to enhance the fine management of typical crop, improve the efficiency and level of agricultural production will play an important role to promote regional economic development. The Internet of things technologies can provide conditions for precise management of crop, but it will also lead to new technical application difficult problems, focusing on the storage and processing problem of the "real time, high frequency, mass, rapid growth" monitoring data of the crop growth environment, this paper put forward a network management and scheduling method for massive data based cloud storage technologies to solve the large data storage and concurrently access bottlenecks of the traditional relational storage, and provided good technical conditions for the typical crop production management.

Keywords: typical crop, growth environment monitoring, cloud storage

1 Introduction

In the growth stage, monitoring of the growth environment of typical crop quantitatively can assist agricultural production managers grasp the crop growth information in time, control the growing conditions in a reasonable manner, so as to achieve the targets of intensive production, increasing yield and efficiency. Currently, through the sensor and the farmland wireless sensor network, the growth information of the crop such as light, temperature, humidity, CO_2 concentration can be monitored and transferred to the remote data centre, inputting them to the production management decision system as the parameters, the agricultural technical personnel can take targeted production measures timely [1].

There are many parameters in the typical crop growth environment monitoring. The environmental monitoring data can be real-time or every few seconds transmitted to the data center, while a data is send in 5 seconds, a sensor node can send 17,280 pieces of data to the server. If there are 1,000 transmit nodes the server will receive 17.28 million pieces data every day. In fact, as the number of typical agricultural product varieties, production field and network nodes increase. The data received will reach far more than this amount, and exceed the capacity of server, it will inevitably lead to many problems such as data volume growth too fast, insufficient database table storage, database system cannot be accessed normally, and etc. [2]. Against the storage and scheduling problems of "real time, high frequency, vast and fast-growing" monitoring data, the mainstream mass data cloud storage technologies were

analysed in this paper, a storage and access method of typical crop growth monitoring data resource was designed and put forward based on massive database solving the bottleneck problem effectively.

2 Distributed cloud storage system

Distributed cloud storage system is one of the core technologies of cloud computing that can integrating a large number of different types of storage devices in the network and provides system functions of data storage and business access through cluster application, grid technology, distributed file systems, and etc. [3].

In this paper, the Internet of things technology was used to monitor the typical crop growth environment, which will inevitably bring large data storage and processing problems [4]. Because the big data has characteristics such as vast and varied, high-speed and mutation [5]. When facing with petabytes of mass data storage requirements, the traditional storage system will create bottlenecks in the capacity expansion and access performance, to resolve this difficult problem, the cloud storage, with its advantage of strong scalability, cost-effective and good fault tolerance can be used, which has been widely recognized in the industry [3].

Currently, the mainstream cloud storage technologies include distributed file system and massive database system. The former is mainly used to store unstructured data, the latter is mainly used to store structured massive data, and part of them can be set up on distributed file system though automatically extending storage space to help the database system automatically expansion and

---

Corresponding author e-mail: wuhr@nercita.org.cn

maintain stable performance. So as to make the growth of the data not be restricted by storage capacity and storage node.

2.1 THE DISTRIBUTED FILE SYSTEM

The Distributed file system is the core of the cloud storage system, the good scalability and fault tolerance features of the cloud storage system depend on the support of the

distributed file system support. The current mainstream distributed file systems include GFS, HDFS, Lustre, fastDFS, PVFS, GPFS, PFS, ceph, TFS and etc. (Table 1), their design thinking is broadly similar, but each one has its own characteristics [3]. Among above systems, the fastDFS is developed by pure c language to achieve the similar lightweight architecture like Google distributed file system, the typical distributed file system was emphatically introduce as follow:

TABLE 1 Mainstream distributed file system

Name	Profile	Features
Google File System (GFS)	A scalable distributed file system, which was mainly used for the large-scale, distributed, intensive data storage.	Runs on a large number of low-cost hardware, provides fault tolerance and high performance services for a large number of users to handle huge amounts of data.
Hadoop Distributed File System (HDFS)	A distributed file system under distributed computing framework Hadoop launched by apache open source organization, providing the underlying support for distributed computing storage.	Taking the error detection and quick recovery as the core goal, having the advantages of high fault tolerance and high throughput, providing the data access model of "written once and repeatedly read" for HDFS applications, and the calculating can be migrated to the node near by the data.
Lustre File System	The distributed file system, launched by Clusterfilesystem Company, is a typical distributed file system based on object storage, the same data files can be divided into several objects stored in different object storage devices, large file I/o operations are assigned to different object storage devices on parallel implementation, so as to realize a large aggregate bandwidth.	Using the expensive and stable server as the file system node, combines the characteristics of the traditional distributed file system and the design concept of traditional shared storage file system. Having the advantages of more effective data management mechanism, the global data sharing, based-on-object-storage, intelligent storage and rapid deployment, so the random access performance is good, but it has none fault tolerance, the storage volume cannot exceed than 1 petabytes, and part of the data will not be accessed when a node failure.
General Parallel File System (GPFS)	A parallel Shared file system launched by IBM which was designed for Linux cluster system drawing on the virtual shared disk technology of IBM Linux cluster system. The compute nodes can parallel access the data of multiple disks in the system at the same time through using of exchange network, and achieve a high I/O bandwidth relying on this access method.	Storing large files on a different disk in a cycle way, reading and writing the small files by merging the operations at the same time, using dynamic election of metadata node to manage the metadata; In addition, the GPFS also has automatic recovery strategy based on log failure node and centralized data locking mechanism.
Parallel File System (PFS)	A distributed file system of Sun company, its main idea is spreading the files across multiple disks and server, and taking the multiple devices as a logical virtual disk to a unified management. PFS can span more than one storage system considering all storage devices at the whole PFS as one part of a virtual disk; When there are multiple nodes access the same file at the same time, PFS can provide access service for these nodes in parallel.	Built on Solaris operating systems, including host nodes, compute nodes, I/O master node and I/O slave nodes. Host node is the entrance of PFS for other systems, only the users who successfully log in to the host node can access the data file within PFS. Compute nodes are primarily used to manage communication system and memory resources of PFS. I/O master nodes are primarily responsible for the directory management and store management of file system, as well as providing services to read and write the stored data files. I/O slave nodes are only used to handle disk read/write operations and allocate of blank block.

These distributed file systems have the functions of spreading the files to the more network storage nodes, and providing the I/O of concurrent access. Considering the factors such as characteristic of agriculture industry, building cost, ease of use and mobility, this paper adopt the HDFS as the distributed file system of the distributed cloud storage system for typical crop production environmental monitoring.

2.2 THE MASSIVE DATABASE SYSTEM

The massive database systems mentioned in this article refers specifically to the NoSQL Database systems corresponding to cloud computing and distributed storage. The NoSQL Database is collectively called the database management system which is different from the traditional

relational, compared to the traditional relational databases which domain the database world for 30 years. The biggest difference is not used SQL (Structured query language) query language; it can store data not in accordance with the fixed table mode, and usually has the characteristics of horizontal scalability. The term "NoSQL" first appeared in 1998, after ten years of development, it evolved into a real technology trend in 2009, pushed by the of big data requirements in the Web 2.0 era, it becomes extremely popular new area and is developing very quickly [7].

Now mainstream NOSQL Database includes Cassandra, Hive, HBase, MongoDB, CouchDB, DynamoDB, Oracle NoSQL, IBM InfoSphere BigInsights, and so on. Among them, Hive, HBase, MongoDB, Oracle NoSQL were used commonly at present, their detailed information was emphatically introduced as follow:

TABLE 2 Mainstream mass-database system (NoSQL databases)

Name	Profile	Features
Hive [8]	Hive is a data warehouse architecture based on Hadoop file system, it uses the MapReduce programming technology, achieving part of the SQL statement, provides the programming interface like SQL.	Hive storage based on Hadoop file system, it does not have a dedicated data storage format itself, also can't establish the index of data, users can very free to organize the tables in Hive, they only need to tell hive the column separators and line separators in the data in order to parse the data when creating a table.
HBase	The Hadoop Database, is a highly reliable, high-performance, column-oriented, scalable distributed storage systems.	Hbase can be built with large-scale structured storage cluster on cheap PC server. It is an open source implementation of Google BigTable using Jave language [7]. HBase is loosely-data stored, its storage mode between the mapping (key/value) and the relational data [8].
MongoDB [7]	An open source product between relational database and NoSql database is one the most abundant NoSql database in function, most like a relational database product.	As a database based on distributed file storage, MongoDB is developed by C++ language, its query languages are very powerful, and it also supports data index.
Oracle NoSQL DB	Oracle NoSQL Database is designed specifically for huge amounts of data management, and it can access the unstructured data, and can be horizontal scalability to hundreds of high availability nodes.	Can be built with Hadoop file system, by means of its own extension mechanism, it can be setup on multiple high-availability node, and has no single point of failure.

These NoSQL database systems can scale horizontally, support high concurrent read and write. They was used not only the traditional relational database model, but also the storage models such as key-value, document type, column, pattern database, XML storage and so on. Most of them are key-value storage [9], the data storage is not bound by ACID theory of traditional relational database, and can expand by freeways, stagger the time, separate partition, store in more than one network node. Taking into account the technological superiority of the horizontal scalability of Oracle NoSQL system itself, bidirectional data exchange with the Hadoop distributed file system and the developing technologies for mass data applications, this article selects the Oracle NoSQL DB as the data storage system of typical crops growth environment monitoring.

3 System design

In order to make cloud storage service technology be applied to the production practice of typical crop growth environment monitoring effectively, this article adopts the sensor acquisition equipment researched and developed independently to get the data, transmitting the data collected from the field to the cloud storage devices in data centre. The cloud storage centre provides uniform data access service of storing the data in the distributed storage unit of the network node automatically, in a logical view, the storage facilities for application are used as a whole storage unit. The horizontal scalability and the parallel processing of mass data are completed by the massive database system and the distributed file system automatically, the overall technical framework of the system is as follows (Figure 1):
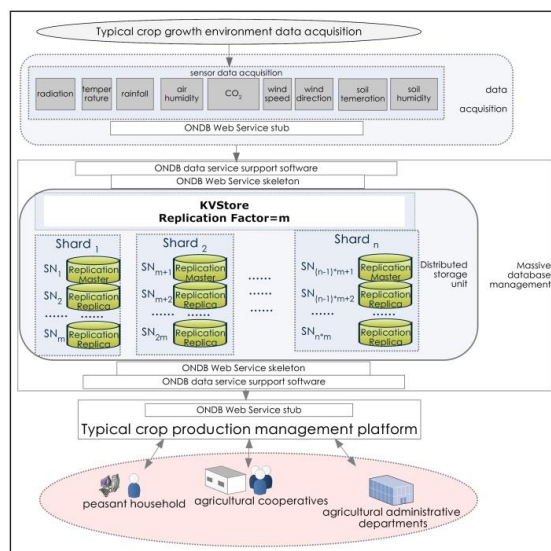
FIGURE 1 Overall cloud storage technology framework for typical crops growth environmental monitoring

3.1 THE DISTRIBUTED FILE SYSTEM

For convenience of system design and implementation, the micro-series of weather station equipment "green cloud grid" which was researched and developed independently was adopted in terms of the field data acquisition and the background data communications can be customized independently.

3.1.1 Acquisition device models

For comprehensive monitoring of typical crop growth environment, the 9 parameters sensor acquisition equipment was chosen to gather the radiation, temperature, humidity, precipitation, carbon dioxide (CO_2) concentration, wind speed, wind direction, soil moisture and soil temperature data through a set of equipment at the same time.

3.1.2 Equipment parameters

The work parameters of sensors acquisition device are as follows (Table 3):

TABLE 3 Working parameters of crop growth monitoring sensor

Sensor	Quantum	Precision	Resolution
Air temperature sensor	[-40,123.8] ℃	±0.5 ℃ (25℃)	0.01 ℃
Soil temperature sensor	[-55, 125]℃	±0.5 ℃ (-10℃ - 85℃)	0.04 ℃
Air humidity sensor	0-100%	±4.5%RH	0.03%RH
Light intensity sensor	0-6 lux or 0-20 lux	5%	0.1 lux
Wind sensor	0-30 m/s	0.1 m /s	0.1 m /s
Wind direction sensor	16 directions	/	1/16 directions
Carbon dioxide(CO$_2$) concentration	0-2000 ppm, 0-5000 ppm, 0-10,000 ppm Optional;	Measurement accuracy: ± 5%; Repeatability: ± 1%.	/
Soil moisture sensor	0-100%	Within 0 ~ 50% (m^3/m^3): ±3% (m^3/m^3)	/
Rainfall sensor	Daily rainfall (0.0 mm - 999.8 mm) Total rainfall (0.0 mm - 9999 mm)	±4 %, ±1 Rainfall counts per hour 0.01 ″- 2.00 ″ (per hour: 0.2 mm and 50.0 mm); ±5 %, ±1 Rainfall counts per hour 2.00 ″- 4.00″ (per hour: 50.0 mm 100.0 mm)	/

3.1.3 Data format

At the scene of the data collection in the crop field, the CAN (control area network) bus was arranged to transmit the data from the sensors to the remote data centre in the form of packet message transmission.

3.2 MASS DATA MANAGEMENT

According to the layer structure of the mass database, it including interface layer, logical model layer, distribution layer and persistence layer [10], in the data centre more than one PC server were chosen to deploy Oracle NoSQL DB. The first step is to set the parameters such as the number of storage Nodes (Store Nodes), the total number of partitions, replicator (Replication Factor), divided area (Shard), and so on. Secondly, the network topology of the cloud storage system must be deployed to make the data storage can scale automatically, can be exchanged with the Hadoop distributed file system, and the data can be stored without being limited by the capacity of the single storage node. Thirdly, research and develop the corresponding Web Service cloud storage middleware of mass data management independently. Fourthly, deploy the middleware skeleton program on the server side to receive the read/write request from clients, and active the middleware itself to complete the physically actions. Finally, deploy the stub program of the middleware both on the PC in CAN bus, and on the PC of typical crop production management platform, so as to realize the integrated management of mass data logically.

4 Key technologies of system implementation

Currently, not like the relation database, the massive databases have neither the data models such as row and line pattern, one-to-one/many-to-many pattern, nor the metadata management mechanism like the relational database management system, without metadata archives, it is difficult to accurately know the exact content stored of the massive database. In this paper, using for reference of the experience of the relational data model building, the Web Service components for modelling, storage and dispatch of the massive data were researched and developed for typical crop growth environment monitoring.

4.1 MASSIVE DATA MODELLING

Most Nosql Databases adopt the key-value pattern or the similar way to store data, so it is difficult to use an intuitive model to describe it. In this paper, according to the business application requirements of the typical crop growth environment monitoring, the data model was still designed as a relational data model to express. Only when the actual program process run, the business logic middleware will execute the different data management mechanism from DBMS system in a much higher efficiency through distributed data process. So this article uses Adobe Fiber Data model to define the model of the typical crop growth monitoring data, the detail model was designed as follows (Table 4):

TABLE 4 Data model of crop growth environmental monitoring data

	The data model diagram	Data items (Field) Chinese name	Data items (Field) English name	Type
The data model		Serial number	dataId	integer
		Node ID	stationId	integer
		Acquisition time	acquisitionTime	datatime
		Illumination (Lux)	radiation	float
		Temperature (ºc)	airTemperature	float
		Humidity (%)	airHumidity	float
		Daily rainfall (mm)	Rainfall	float
		Total rainfall (mm)	allRainfall	float
		CO2 Concentration (PPM)	CO2	float
		Wind speed (m/s)	windSpeed	float
		Wind direction	windDirection	string
		Soil water content (%)	soilHumidity	float
		Soil temperature (ºc)	soilTemperature	float

4.2 MASSIVE DATA PERSISTENCE

In order to facilitate the migration from DBMS-orient application to mass-database-orient applications, decrease the difficulty of the application development, and can use the JPQL grammar similar to relational database to access massive database data in general case. This paper adopted the data persistence framework to define a Java object model corresponding to the data model in section 4.1 at first, and wrote a web service entity rwSensorData as follow (Table 5):

TABLE 5 Crop growth environment monitoring data Java Object model

	Java Object Model(DTO)	Key programs (pseudocode)

Java Object Model		import javax.persistence.EntityManager;
		import javax.persistence.EntityManagerFactory;
		import javax.persistence.Persistence;
		import javax.persistence.Query
		import javax.persistence.*;
		import org.eclipse.persistence.nosql.annotations.NoSql;
		import org.eclipse.persistence.nosql.annotations.DataFormatType;
		@Entity
		@NoSql(dataFormat=DataFormatType.MAPPED)
		public class SensorData {
		    @Id
		    @GeneratedValue
		    private double dataId;
		        ……..
		    @Basic
		        private Double windSpeed;
		    public double getDataId() {
		        return dataId;
		    }
		    public void setDataId(double dataId) {
		        this.dataId = dataId;
		    }
		    ……
		    public Double getWindSpeed() {
		        return windSpeed;
		    }
		    public void setWindSpeed(Double windSpeed) {
		        this.windSpeed = windSpeed;
		    }
		}
		}

From the formal point of view, the model above was similar to the Java object model of relational database. The differences mainly consists of the note symbols started with "@" which inform the data persistence framework to perform the corresponding operation in Oracle NoSQL database. So it is convenient to integrate it with traditional application and then based on the data persistence management object (EntityManager). The middleware rwSensorData can use Sensordata entity object as data transfer unit to achieve the basic interfaces of massive data such as read, write, delete, and modify operation.

4.2.1 Writing massive data

The monitoring data is stored into massive database in decentralized order and by parameters. As soon as data transited from sensors to server-side, the framework program (skeleton) of entity class rwSensorData will call the set method of SensorData, pass the data value from sensors to SensorData object at first, then call the persist method of data persistence framework to store the data into the massive databases taking the SensorData entity object as unit. Each of the "field" is individually deposited, the entire process of writing data is completed in a concurrent and distributed environment, and therefore there is no data congestion.

4.2.2 Reading massive data

In application layer, using JPQL persistence language, which is consistent with the relational database can read the massive data from Oracle NoSQL. The underlying operation is completed by data persistence framework through calling Oracle ONDB API for data reading.

4.2.3 Other massive data operation

Similarly, based on the data persistence management object (EntityManager) of massive data persistence framework, the other operations such as modifying and deleting of massive data can also be finished, through calling the find method of the EntityManager. The object data can be located for modification by assign new value, or deleting the data by call the remove method.

5 Application effectiveness analysis

Finally, in this paper, the "tomcat 6.0.14+BlazeDS" technology architecture was used under the JavaEE environment, the multi-node storage, multiple copies of fault-tolerant and free extension of mass data cloud storage system was deployed based on Oracle NoSQL DB, and the typical crop production management decision platform prototype was researched and developed to gather the monitoring data from crop planting field and provide production guidance services for the users at the other network nodes.

5.1 CLOUD STORAGE SYSTEM DEPLOYMENT AND OPERATING RESULTS

In this paper, the cloud storage system was designed according to the persistence layer, logic model layer, distribution layer and interface layer, the specific deployment way and function are introduced as follows:

TABLE 6 Cloud storage system deployment

Component	Deployment	Function
Interface layer	Deployed the rwSensorData in Spring + BlazeDS framework	Provides resource pooling, connection pools
Distribution layer	Deploying 5 network nodes of Oracle NoSQL Nodes and Hadoop nodes, (1 master node and 4 slave nodes, the initial configuration data storage capacity of each node is 20T)	Automatically scale data in multiple node, maintaining the node data autonomy
Logical model	EclipseLink JPA Java Object -Map	Keeping the same modeling method with the relational database, the difference only exits in physical storage in the underlying interface.
Persistence layer	EclipseLink 2.4.1 JPA for Oracle NoSQL	Shielding the differences in the underlying interface of data provider, the application layer provides the consistent style of data persistence interfaces with the relational database.

According to the above design, when the monitoring data is received via PC or SMS module from sensor node to the cloud storage system, an idle instance from the interface pool (instance) at the server-side will be assigned to complete data storage operation. The distribution layer will store the receive decentralized data into different network nodes in real time, the data that has the same key value will be automatically stored in a partition (Figure 2), and the database system will store the copies of data according to pre-configured number of replication factor

of the cloud storage system at the same time. The distributed storage was autocompleted by NoSQL and Hadoop due to the data access operation is responded by interface pool. The data operation will be completed in instantaneous, and there will not occur server goes down, slow response and stopped responding phenomenon.

Use data query interface of Oracle NoSQL, the data stored in a massive database can be seen as the following Figure 2:

```
 1 [CUSTOMER, D5BA62AE-3356-4C0A-AE04-48DC6D07B893, ID]D5BA62AE-3356-4C0A-AE04-48DC6D07B893
 2 [CUSTOMER, D5BA62AE-3356-4C0A-AE04-48DC6D07B893, NAME]AMCE
 3 [KJCDIGITALMARKET, 19545, Boneless sheep hind legs]19545|Boneless sheep hind legs|18|18.5|19|General||2010-11-06 00:00:00|http://www.xinfadi.co
 4 [KJCDIGITALMARKET, 19557, Beef heart]19557|Beef heart|5|5|5|General||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.php?mid=14amp;
 5 [CUSTOMER, 481A8827-3DF2-41D8-8CD4-D850CD4C34D2, ID]481A8827-3DF2-41D8-8CD4-D850CD4C34D2
 6 [CUSTOMER, 481A8827-3DF2-41D8-8CD4-D850CD4C34D2, NAME]Smith
 7 [KJCDIGITALMARKET, 19554, Steak]19554|Steak|12|12.25|12.5|General||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.php?mid=14amp;
 8 [KJCDIGITALMARKET, 19539, Sheep ribs]19539|Sheep ribs|13|13.5|14|||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.php?mid=14amp;pa
 9 [KJCDIGITALMARKET, 19542, Mutton kidney]19542|Mutton kidney|20|20|20|||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.php?mid=14
10 [KJCDIGITALMARKET, 19543, Mutton kidney]19543|Mutton kidney|17|17.5|18|General||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.ph
11 [KJCDIGITALMARKET, 19546, Boneless sheep forelegs]19546|Boneless sheep forelegs|17|17.5|18|General||2010-11-06 00:00:00|http://www.xinfadi.com
12 [KJCDIGITALMARKET, 19550, Shelter]19550|Shelter|12|12|12|General||2010-11-06 00:00:00|http://www.xinfadi.com.cn/do/listform.php?mid=14amp;pa
13 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, DATAID]1
14 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, STATIONID]1
15 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, ACQUISITIONTIME]2012-05-03 07:30:00
16 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, RADIATION]1700
17 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, AIRTEMPERATURE]15
18 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, AIRHUMIDITY]10
19 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, RAINFALL]0
20 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, CO2]26.1
21 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, WINDSPEED]21.5
22 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, WINDDIRECTION]SSW
23 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, SOILHUMIDITY]30
24 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, SOILTEMPERATURE]20
25 [SENSORDATA, FA18AE16-A43F-4437-861B-DF1990976782, VERSION]1
```

FIGURE 2 Data stored in the massive database system

In Figure 2, the data whose primary key is "SENSORDATA" can be think as its database table name is "sensordata" like in relational database, the minor keys "FA18AE16-A43F-4437-861B-DF1990976782" can be think as the row recorder number in the relational database, the same batch of data of radiation, temperature, humidity, rainfall, carbon dioxide (CO_2), soil moisture, soil density, wind speed, wind direction and soil temperature have the same minor keys (for example, "FA18AE16-A43F-4437-861B-DF1990976782") and the same version number. The name of the last minor key can be interpreted as the field names in the relational database, and key value followed by "field" is its specific data values. Comparison with the relational database, the data in massive database is organized as "one by one grain", it hasn't fixed format, and

hasn't constraints of ACID theory in the relational database, the style is very similar to the Hash table in the Java language, so as to the performance efficiency is very high, and it is easier to extend in storage.

5.2 PERFORMANCE EFFICIENCY ANALYSIS

In order to understand the cloud storage system efficiency intuitively, in this paper, a simple test environment of "64 Bits PC, 8G Memory, 3×3 Storage nodes, 10 concurrent visits" was deployed under the experimental conditions. The response time was observed respectively on the amount of data request of 20-160 records by the benchmarks test of cloud service, the test results are as follows (Figure 3):
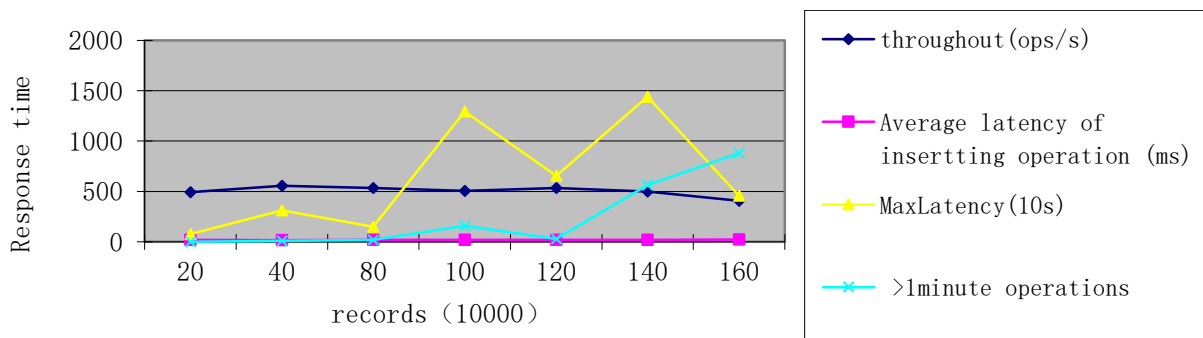
FIGURE 3 Massive database and relational database data access response time comparison chart

From above figure test case, to keep the high concurrent access throughout performance of the mass data as a whole, the mass database system stored the data "field by field" according to the data column. If the topology was designed reasonably, it always keep comparison smooth of data operation response performance. The average delay of insetting data insert shows slight fluctuations along a horizontal line, as the data scale of increased, the data throughput performance will decline, and the maximum delay time and the operations over 1 minute will increase, but all the data will

be stored at some point later than the request time eventually.

6 Conclusion and discussion

Against the technical problems of traditional database on massive data processing, the cloud storage technology for typical crop growth monitoring data was researched in this paper. The data storage method was put forward based on Oracle NoSQL massive database, which can transmit and stored the huge data of radiation, temperature, humidity, rainfall, carbon dioxide (CO_2) concentration, soil water,

soil temperature, wind speed, wind direction from sensor network nodes to the cloud storage centres of typical crop production management platform respectively. Moreover, solved the efficient access issues of the "real-time, high-frequency transmissions, the fast-growing" massive monitoring data, the main findings was made as following:

1) The common clouds distributed storage systems was invested and analysed. The cloud storage method suitable for typical crop growth monitoring mass data was put forward using the storage framework of combination the Oracle NoSQL and Hadoop FS, and solved the bottlenecks of concurrent access of massive data by means of the distributed storage mechanism for the mass data among multiple network node.

2) In order to reduce the difficulty of migrating legacy applications to the massive database system application target, the cloud service framework for mass data access was raised to be consistent with the physically distributed storage system. The cloud service interface for mass data access was researched and developed to provide the

technical support for the deep application of typical crop growth monitoring data resource.

3) The execution efficient of the cloud storage technology for typical crop growth monitoring data was analysed. Under the condition of the same data scale application, the mass data storage system has the characteristics of automatic expansion of the capacity and distributed concurrency computing in real-time, compared with the traditional relational database systems, it has more greater data throughput performance, more reliable operation, and can meet the challenges of mass data from typical crop growth environment monitoring.

Acknowledgments

References

[1] Zeng X H, Wang Q 2011 General situation of agricultural information technology and network *China Agricultural Science Bulletin* **27**(8) 468-73

[2] Jia X F 2011 Research on the strategy for real-time information exchange of the mass data based on the Internet of things *E-Government* **4** 16-20

[3] Deng J G, Pan X H, Yuan H Q 2012 Research on cloud storage and its distributed file system *Journal of Dongguan University of Technology* **19**(5) 41-6 *(in Chinese)*

[4] Ning H S, Xu Q Y 2010 Research on global internet of things' developments and it's instruction in China *Acta Electronica Sinica* **38**(11) 2590-9

[5] http://www.enet.com.cn/cio/zhuanti/2012/bigdata

[6] Zou C H 2011 Research of adaptive local hyperplane K-nearest neighbour classification algorithm based on hadoop platform *South China University of Technology (in Chinese)*

[7] Yu X 2012 NoSQL: a rise in the great wave of big data *Network World* **4** 1-5

[8] Lu J H 2012 The Practice of Hadoop *Beijing: Machinery Industry Press (in Chinese)*

[9] Zhang H Q 2011 Relational database and NoSQL database, Computer Knowledge and Technology **7**(20) 4802-4

[10] Shen S 2012 Research on NoSQL Database Technology and Application *Nanjing: Nanjing University of Information Science and Technology (in Chinese)*

Authors

Yuansheng Wang, born on July 7, 1973, Wuhan, Hubei province, China

Current position, grades: associate professor of Beijing Research Centre for Information Technology in Agriculture.
University study: Ph.D. degree in cartography and geographic information systems, Beijing Forest University in 2008.
Scientific interest: agricultural and rural information resource integration, agricultural eco-environment, forest resources and forest fire prevention, agricultural LBS.
Publications: more than 10 journal papers since 1997
Experience: mainly engaged in the study of agricultural information technology since 2003.

Huarui Wu, born on July 21, 1975, Liaocheng, Shandong province, China

Current position, grades: professor of Beijing Research Centre for Information Technology in Agriculture.
University study: Ph.D. degree in computer applications, Beijing University of Technology in 2010.
Scientific interest: agricultural and rural information resource integration, agricultural Intelligent Systems, wireless sensor network technology.
Publications: more than 30 journal papers and co-authored several government reports since 2000.
Experience: mainly engaged in the study of agricultural information technology since 2001.

Lei Luo, born on August 10, 1987, Baoding, Hebei province, China

Current position, grades: engineer, Beijing Research Center for Information Technology in Agriculture.
University study: Master degree in computer application, Hangzhou Dianzi University, Hangzhou, China, in 2013.
Scientific interest: application of information technology in agriculture.
Experience: agricultural information technology engineer since 2013.

Weihong Ma, born on February 18, 1987, Xingtai, Hebei province, China

Current position, grades: Beijing Research Centre for Information Technology in Agriculture
University study: Master's degree in mechanical and electronic engineering, Beijing University of Technology in 2011.
Scientific interest: agricultural and rural information resource integration.
Experience: agricultural information technology engineer since 2011.

Kun Liu, born on December 18, 1990, Taian, Shandong province, China

Current position, grades: Software engineering student, Beijing Research Center for Information Technology in Agriculture.
University study: Master's degree, Shandong Agricultural University, Shandong, in 2014.
Scientific interest: research and application of information technology in agriculture.

Computer and Information Technologies

Image adaptive filtering based on the improved Alpha-trimmed mean algorithm

Wangcheng Cao*, Lin Zheng, Kuiling Dong

School of Engineering, Mudanjiang Normal University, Mudanjiang 157011, Heilongjiang, China

Abstract

Filtering is an important research direction of image processing. In view of the characteristics of image noise, an adaptive image filtering algorithm is proposed based on the improved Alpha-trimmed mean algorithm. The algorithm dynamically selects parameter d of Alpha-trimmed mean algorithm through the calculation of the pixel correlation within the filter window, so that the algorithm can filter accordingly with the degree of noise. The experiment results show that the filtering effect of the proposed algorithm is excellent for the images corrupted by Gaussian noise, salt and pepper noise or mixed noise, and this algorithm is also capable of maintaining the detailed information of the original image.

Keywords: image filtering, alpha-trimmed mean filter, neighbourhood correlation

1 Introduction

Image is a kind of file, which is easier to be seen and used in people's daily life. So the studies for image processing method have been carried out for many years and have achieved good results in many areas [1, 2]. However, whether the traditional imaging devices or modern digital imaging equipments, most of them worked based on the principle of transmission and refraction of the light. So the imaging process easy to be affected by the natural or human factors, resulting the image contains much noise and not clear enough [3, 4]. In addition, modern digital cameras and other electronic equipments will be involved in the use of related integrated circuits which also exist some features that affect the imaging process, such as drift, edge nonlinearity and so on [5]. These factors can also cause the image contains noise more or less, so the image filtering has become a research focus.

The main purpose of the image filtering is to remove the noise and improve image quality, which directly affect the subsequent analysis of the image processing. So we not only consider the noise filtering ability of an algorithm, but also to try to keep details of the original image when carried on the image filtering. For all the image filtering methods, linear filtering is studied much earlier, mainly includes Gaussian filter, mean filter, Wiener filter, etc. [6,7]. The advantage of these methods is very effective for Gaussian noise, but cannot filter the impulse noise, and will also cause the loss of image details. To compensate for the defect of the linear filtering algorithm, then many scholars studied another form of filtering algorithm—nonlinear filtering algorithm, mainly includes median filter, statistical filter, morphological filter, etc. [8-10]. Such filtering algorithm can better maintain the details of

the image and can filter the impulse noise effectively, but these algorithms emphases on the ability of maintaining the details so much, so cannot filter Gaussian noise with the normal distribution. Generally, the image noise includes many kinds of noises, in this case, whether to use a linear filter or non-linear filter cannot yield satisfactory results.

To solve these problems researchers have proposed an Alpha-trimmed mean filtering algorithm that fully makes use of the advantages of linear filters and nonlinear filters. This algorithm solves the problem of image filtering to a certain extent, especially in the case of mixed noise exists. However, this algorithm also has some defects. Firstly, the size of the filter window is fixed, some useful information is not fully used to deal with the problem that the noise of the image with different intensity. Secondly, the parameter *d* of the algorithm is determined according to the human experiences, which cannot be changed adaptively with different circumstances. These factors limit the promotion and using of the algorithm. Therefore, in this paper, the parameter *d* is selected dynamically by the noise intensity within the filter window and the noise intensity is estimated by the mean and standard deviation of the pixel neighbourhood correlation. The innovation of this paper is to make the traditional Alpha-trimmed mean filtering algorithm has the adaptive ability. The simulation results also proved the correctness and effectiveness of the filtering algorithm proposed in this paper.

2 The Alpha-trimmed mean filtering algorithm

Assume that the centre pixel within the filter window is $f(x, y)$, where x and y are pixel coordinates, $f(x, y)$ is grey value. The size of the window is $n \times n$, all the pixels

within the window represented by $g_p(s,t)$ $p \in n \times n$, then removes $d/2$ max pixels and $d/2$ min pixels, the left $n^2 - d$ pixels represented by $g_c(s,t) c \in n^2 - d$. Finally, the Alpha-trimmed mean filtering algorithm is given as follows, see Equation (1):

$$\hat{f}(x,y) = \frac{1}{n^2 - d} \sum g_c(s,t),$$ (1)

where, $d \in \left[0, n^2 - 1\right]$. When $d = 0$, the Alpha-trimmed mean filtering algorithm equal to the mean filter, while it equal to median filter when $d = (n^2 - 1)/2$.

It can be seen from the Equation (1) that the performance of the algorithm affected largely by the parameter d. When the type of image noise is Gaussian noise, the parameter d should not be too large; while when the noise is salt and pepper noise, the parameter d should not be too small. The situation is even more complicated when the image contains some different noises. Thus, although the principle of the algorithm is relatively good, the satisfied filtering result cannot be got when the parameter d is not suitable.

3 The neighbourhood correlation

The characteristic of neighbourhood information relation exists among all the adjacent pixels, so the difference between adjacent pixels can reflect their degree of association. This characteristic is used to calculate the neighbourhood correlation within the filter window in this paper. For the convenience of description, the 3×3 window is as an example, please seeing Figure 1.

1	2	3
4	5	6
7	8	9

FIGURE 1 The 3×3 filter window

Firstly, calculates the neighbourhood correlation of each pixel. For example, the pixel 1 is calculated as follows, see Equation (2):

$$r_1 = \frac{\sum_{j=2,4,5} \left| f_1(x,y) - f_j(x,y) \right|}{3}.$$ (2)

The neighbourhood correlation of other pixels r_i, $i = 2, \cdots, n$ calculated follows this method. Next, calculate the mean of neighbourhood correlation within the window by the Equation (3):

$$r_{avg} = \frac{\sum_{i=1}^{n} r_i}{n}.$$ (3)

Next, calculate the standard deviation of the neighbourhood correlation by the Equation (4):

$$\sigma_r = \sqrt{\frac{\sum_{i=1}^{n} (r_i - r_{avg})^2}{n-1}}.$$ (4)

Now, we have got all the needed parameters of pixel neighbourhood correlation.

4 Implementation steps of improved Alpha-trimmed mean filtering algorithm

As described in Section 2, the Alpha-trimmed mean filtering algorithm has some flaws, i.e. the parameter d and the size of filter window are fixed value and cannot be adaptive changes according to the situation. The improved method presented in this paper is: the intensity of noise within the filter window is calculated firstly, and then the value of the parameter d is determined by it. While the intensity of noise is estimated mainly through calculates the neighbourhood correlation of each pixel within the filter window. The specific process is: first calculates the neighbourhood correlation of each pixel within the filter window, and then calculates the mean and standard deviation of the neighbourhood correlation, finally, using the 3σ rule to count the amount of noise, that is, the pixel is noise when its neighbourhood correlation greater than $r_{avg} + 3\sigma_r$. The size of the filter window is adjusted by whether $n^2 - d$ greater than 1. Based on the principle of the improved method described above, the specific implementation steps are as follows:

Step 1: selects an initial $n \times n$ filter window, n is odd and $n \geq 3$;

Step 2: calculates the neighbourhood correlation of each pixel according to the Equation (2), and then calculates its mean and standard deviation by the Equations (3) and (4) respectively;

Step 3: calculates the parameter d according to the amount of pixel whose neighbourhood correlation greater than $r_{avg} + 3\sigma_r$;

Step 4: go to the next step if $n^2 - d > 1$, otherwise, enlarge the filter window and go to step 2;

Step 5: calculates the filtering output result by the Equation (1).

According to the improved method above, the parameter d and the size of the filter window of the Alpha-trimmed mean filtering algorithm can be selected adaptively by the intensity of noise, so the filtering effect is better and the efficiency of the algorithm is higher.

5 Experiment analysis

In order to test and verify the efficiency of the algorithm proposed in this paper, the software Matlab and its image library are used to carry on some simulation experiments and the filtering performance will be compared to the mean

Cao Wangcheng, Zheng Lin, Dong Kuiling

filter and median filer, which are very representative. Now, the image Lena in the image library is selected as the test example, and the test process includes 3 cases:

1) we will add Gaussian noise in the image Lena with different intensity;

2) we will add salt and pepper noise in the image Lena with different intensity;

3) we will add mixed noise in the image Lena with different proportion. The criterion of objective image fidelity is used to evaluate the effect on image filtering, which includes the following two expressions:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (S_{ij} - Y_{ij})^2 , \qquad (5)$$

$$PSNR = 10 \lg \frac{255^2}{MSE} , \qquad (6)$$

where, *M, N* are the height and width of the image respectively, S_{ij} is the original image, Y_{ij} is the image after filtering. We can see from the Equation (5) that the smaller the value of *MSE*, the better image fidelity, that is, the noise is filtered effectively. The similar conclusion can be got from Equation (6), while the difference is the larger the value of *PSNR*, the better filtering result, image after filtered closer to the original image.

The filtering results are shown in Figures 2-7, where, the intensity range of Gaussian noise and salt and pepper noise is 0.02~0.4.
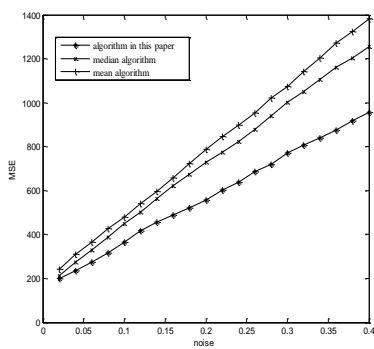
FIGURE 2 The *MSE* result of Gaussian noise with different intensity
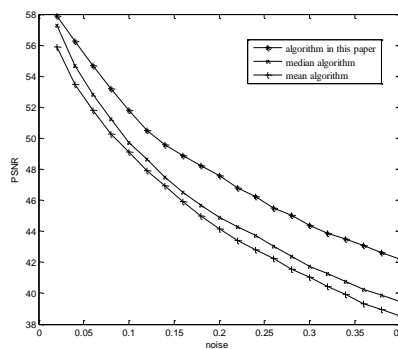
FIGURE 3 The *PSNR* result of Gaussian noise with different intensity
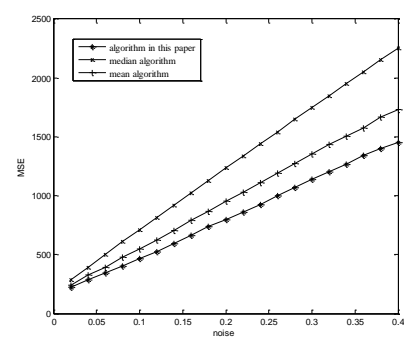
FIGURE 4 The *MSE* result of salt and pepper noise with different intensity
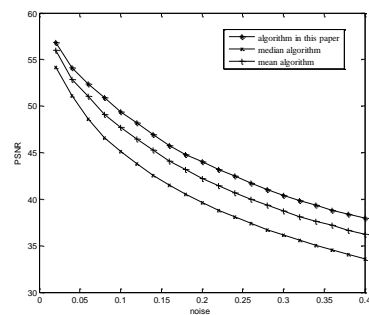
FIGURE 5 The *PSNR* result of salt and pepper noise with different intensity
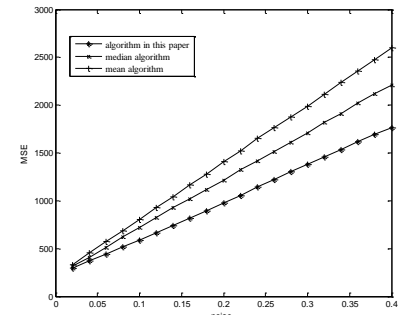
FIGURE 6 The *MSE* result of mixed noise with different intensity
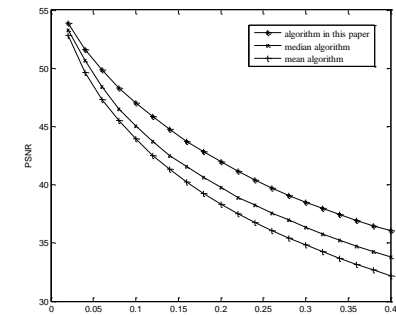
FIGURE 7 The *PSNR* result of mixed noise with different intensity

It can be seen from Figures 2 to 7 that the mean algorithm, median algorithm and the algorithm in this paper all could reduce the *MSE* of the image in varying extents, and enhance the *PSNR* simultaneously. We can easily find that, however, the filtering results of the algorithm in this paper for Gaussian noise and salt and pepper noise with different intensity, and for mixed noise with different proportion all so good, and can remain more local details of the image. Moreover, the filtering results of the traditional mean filter and median filter for mixed noise are poor, where the mean filter is more sensitive to salt and pepper noise, while the median filter is more effective to the Gaussian noise. Therefore, the overall performance of the proposed filter algorithm is better.

In order to further judge the filtering effect of the algorithm presented intuitively, the image filtering output

also given in this paper. The result is shown in Figure 8, where the noise is a mixture of the Gaussian noise with intensity 0.01 and salt and pepper noise with intensity 0.1.
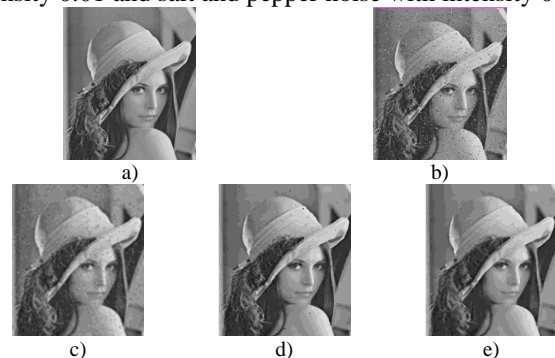
FIGURE 8 The filtering result of various algorithms

As seen from the visual effects in Figure 8, the proposed method shows better filtering performance, enabling a sharper edge of the original image. This fully shows that the true information of the image is better protected by the algorithm in this paper, the ability to protect the edge of the image also to be improved, and can filter the noise effectively. In short, the proposed algorithm is much better, whether the subjective or objective analysis.

6 Conclusions

The image is likely to be mixed with the noise when imaging, transmission, processing, display and other operation by various reasons, which can affect the clarity of the image, and is not propitious for the subsequent processing. To solve this problem, this in-depth comparison of the advantages and disadvantages of the traditional linear filtering method and nonlinear filtering method is carried on in this paper, and then we propose a new filtering method. This method introduces the concept of neighbourhood correlation of pixels whose mean and standard deviation is used to select the parameter d of the traditional Alpha-trimmed mean filtering algorithm dynamically, simultaneously, this method can adjust the size of the filter window according to the situation. The improved method not only retains the advantages of the traditional Alpha-trimmed mean filtering algorithm, but it has adaptive characteristics. Simulation results show that the noise filter effect of the improved algorithm is better for Gaussian noise, salt and pepper noise and mixed noise, a significant improvement over the existing median filtering and mean filtering algorithm. The capacity of the filtering and image details keeping are both good, while the efficiency of the algorithm is enhanced greatly because of the adaptive feature, which is very suitable for practical application.

References

[1] Takahashi N, Fujita K, Shibata T 2009 *IEEE Transactions on circuits and systems I regular papers* **56**(11) 2384-92
[2] Gao Y, Peng J, Luo H, Keim D A, Kianping F 2009 *IEEE Transactions on circuits and systems for video technology* **19**(12) 1851-65
[3] He K, Sun J, Tang X 2013 Guided Image Filtering *IEEE Transactions on pattern analysis and machine intelligence* **35**(6) 1397-409
[4] Milanfar P 2013 *IEEE signal processing magazine* **30**(1) 106-28
[5] Li Y, Sun J, Luo H 2014 A neuro-fuzzy network based impulse noise filtering for gray scale images *Neurocomputing* 127 190-9
[6] Dolui S, Kuurstra A, Patarroyo S I C, Michailovich O V 2013 A new similarity measure for non-local means filtering of MRI images Original Research Article *Journal of visual communication and image representation* **24**(7) 1040-54

[7] Maruyama T, Yamamoto H 2011 *IEEE Transactions on image processing* **5**(5) 457-65
[8] Yang C-C 2013 Improving the overshooting of a sharpened image by employing nonlinear transfer functions in the mask-filtering approach *Optik-international journal for light and electron optics* **124**(17) 2784-6
[9] Zhang Y-Q, Ding Y, Liu J, Guo Z 2013 *IEEE Transactions on image processing* **7**(3) 270-9
[10] Rahman M M, Antani S K, Thoma GR 2011 *IEEE Transactions on information technology in biomedicine* **15**(4) 640-6
[11] Oten R, De Table
[12] ueiredo, Rui J P 2004 *IEEE Transactions on image processing* **13**(5) 627-39
[13] Ahmed F, Das S 2013 Removal of high density salt-and-pepper noise in images with an iterative adaptive fuzzy filter using alpha-trimmed mean *IEEE Transactions on fuzzy systems* (99) 1-2

Authors

Wangcheng Cao, born on December 29, 1980, China

Current position: lecturer at School of Engineering, Mudanjiang Normal University, China.
University studies: Master of Engineering degree in Technology of Computer Application from Harbin University of Science and Technology, China in 2007.
Scientific interest: microcontroller and embedded system development.

Lin Zheng, born on December 8, 1984, China

Current position: lecturer at School of Engineering, Mudanjiang Normal University, China.
University studies: Master of Engineering degree in Technology of Computer Application from Harbin University of Science and Technology, China in 2009.
Scientific interest: business intelligence, network and information security.

Kuiling Dong, born on September 4, 1981, China

Current position: lecturer at School of Engineering, Mudanjiang Normal University, China.
University studies: Master of Engineering degree in Computer Science and Technology from Harbin University of Science and Technology, China in 2007.
Scientific interest: network and information security.

Optimal interpolation data assimilation of surface currents by utilizing pseudo measurement with Monte Carlo simulation

Lei Ren[*], Stephen Nash, Michael Hartnett

[1]*Department of Civil Engineering, National Univerisity of Ireland, Galway, University Road, Galway, Ireland*

[2]*Ryan Institute, Galway, Ireland*

Abstract

Optimal Interpolation (OI) data assimilation is a technique to combine available observations with background states to improve prediction states. In this research, pseudo measurement of surface currents generated by adding noise with Monte Carlo simulation is used to update the background states with optimal interpolation. The core of Optimal Interpolation data assimilation is the definition of background error covariance, which determines to what extent the model background states will be corrected to match the observations. The background error covariance is computed before the data assimilation process. The model background errors are calculated from the mean over a short time interval ten minutes. A series of sensitivity tests with Optimal Interpolation are done by calculating Root Mean Square Error (RMSE) to decide the appropriate parameters. The improvement of Optimal Interpolation at reference points is measured in Taylor diagrams, and the surface current maps of test domain show the effectiveness of Optimal Interpolation.

Keywords: Optimal Interpolation, data assimilation, Monte Carlo, pseudo measurement, background error covariance

1 Introduction

Data assimilation is a technique to improve the modelling prediction ability by blending available measurement information with the background states. In general, there are two kinds of data assimilation algorithms: sequential and variational data assimilation (Robinson and Lermusiaux, 2000, Moore, Arango, Broquet, Powell, Weaver and Zavala-Garay, 2011, Ma, Zheng, Zhong and Zou, 2014). The analysis equation of the former is expressed by the linear combination between background and measurement states; the latter algorithm is generally derived from an objective function measuring the distance between observed states and background states (Zaron, 2009, Dong and Xue, 2012). In our current work, sequential Optimal Interpolation data assimilation scheme is undertaken to update the model background states.

For sequential Optimal Interpolation data assimilation schemes, innovation of calculating the background error covariance results in a variety of methods, such as Optimal Interpolation (Gu, Woo and Kim, 2011, Rienecker, 1991), Ensemble Optimal Interpolation (EnOI) (Oke, Brassington, Griffin and Schiller, 2010, Counillon and Bertino, 2009). Due to the inexpensiveness and flexibility of Optimal Interpolation data assimilation algorithm, it is becoming a popular data assimilation approach in oceanography (Counillon and Bertino, 2009, Oke, Brassington, Griffin and Schiller, 2008, Srinivasan, Chassignet, Bertino, Brankart, Brasseur, Chin, Counillon, Cummings, Mariano, Smedstad and Thacker, 2011).

Optimal Interpolation and Ensemble Optimal Interpolation data assimilation had been applied in some operational oceanic hydrodynamic prediction systems (Oke, Brassington, Griffin and Schiller, 2010, Oke, Brassington, Griffin and Schiller, 2008, Carton and Giese, 2008). In this paper, Optimal Interpolation method is used to update the surface velocity components by using pseudo measurement generated with Monte Carlo simulation (Doucet, de Freitas and Gordon, 2001). In order to clearly analyse Optimal Interpolation data assimilation process, a small test area is defined as our data assimilation domain, the purpose of this work is to develop an Optimal Interpolation data assimilation system for coastal areas and assess the improvement of Optimal Interpolation data assimilation.

An outline of this paper is as follows: Section 2 and 3 describe the three dimensional numerical modelling and generation process of pseudo measurement. Section 4 presents Optimal Interpolation data assimilation schemes. Results of Optimal Interpolation data assimilation is presented in Section 5, followed by conclusions of Optimal Interpolation data assimilation in Section 6.

2 Numerical modelling

The Environmental Fluid Dynamics Code (EFDC) is applied to simulate the dynamic process of Galway Bay, which is located on west coast of Ireland. The numerical model EFDC solves the three-dimensional, vertically hydrostatic, free surface, turbulent averaged equations of

---
[*] *Corresponding author* e-mail: leirencomeon@gmail.com

motions for a variable density fluid. The module uses a sigma vertical coordinate and curvilinear, orthogonal horizontal coordinates. There are 380×241 grids in the rectangular simulation domain of model, the grid resolution is 150 metres, the physical domain is from (-9.71891E, 52.97371N) (left at the bottom) to (-8.87716E, 53.03773N) (right on the top), following picture shows the modelling area and data assimilation domain. The basic research area is Galway bay, a square domain with 961 wet cells is defined as our data assimilation domain, real dimension is a 4.65km×4.65km square area.
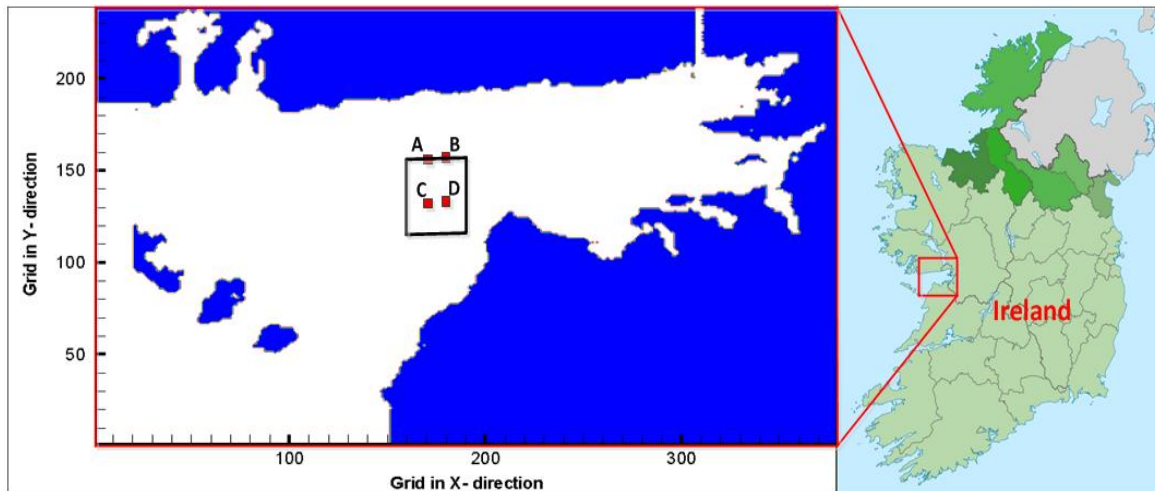
FIGURE 1 Research area and data assimilation test domain

The meteorological data (temperature, rain, solar radiation and relative humidity etc.) are obtained from the weather station located at National University of Ireland, Galway (http://weather.nuigalway.ie/). The river inflow of River Corrib was got from the Office of Public Works (http://www.opw.ie/hydro/). And the tidal information is generated from Oregon State University Tidal Prediction Software (OTPS), which provides tidal information on the western and southern open boundaries. In order to illustrate the Optimal Interpolation data assimilation process in detail, there are three main simulations performed: the Free run, which is initialised with no data assimilation for seven days (01/10/2011-07/10/2011), its output is applied to compute the background error covariance; the Assimilation run, which is initialised for seven days (14/10/2011-20/10/2011), but the surface velocity components are updated by utilising the pseudo measurement during the last four days; the Control run is the same as the Assimilation run, but with no data assimilation, which is regarded as the standard reference state, its output of surface velocity components are used to generate pseudo measurement by adding normal distribution noise.

3 Pseudo measurement

In order to update the background states in numerical model, pseudo measurement is generated by adding normal distribution noise to the output from a model run with Monte Carlo simulation, Monte Carlo simulation is undertaken to yield pseudo measurement based on the basic hydrodynamic trend(Doucet, de Freitas and Gordon, 2001). In order to clarify the difference between original model results and generated pseudo measurement, the noise is unbiased, the standard deviation of added noise is 20% of the maximum absolute difference of velocity components during model stable phase, the value of standard deviation are 4 cm/s and 3 cm/s for surface velocity components (u and v) separately. There are three obvious advantages by using this kind pseudo measurement: firstly, the generated pseudo measurement is based on the output from Control run, the general trend of generated pseudo measurement still follows the basic dynamical process; secondly, the value of noise can be controlled, which means that the sensitivity test of Optimal Interpolation data assimilation can be more accurately assessed; thirdly, the generation process of pseudo measurement is based on the background states field from the Control run and the noise field has the same structure of background states, so the yielded artificial measurement field matches well with the model state field.

4 Optimal interpolation data assimilation

Optimal Interpolation data assimilation combines observation states with model background states to obtain better prediction results, here, the background states are updated by utilising pseudo measurement.

The analysis equation of Optimal Interpolation is a linear combination of background states and measurement states, the optimal weight factor (Kalman gain) is derived by minimizing the analysis error covariance, the analysis equation could be expressed as follows (Kalnay, 2002):

$$x^a = x^b + K(y^0 - Hx^b) \, , \tag{1}$$

241

where x^a is the analysis state, x^b is the forecast or background state, K is the Kalman gain, H is the measurement operator, y^0 is the observation state.

The state in our research is surface velocity components u and v, which could be given as:

$$x = (u, v)^T . \tag{2}$$

The Kalman gain is obtained in the following formula by minimizing the analysis error covariance (Kalnay, 2002):

$$K = P^b H^T (HP^b H^T + R)^{-1}, \tag{3}$$

where P^b is the background error covariance, R is the measurement error covariance

The steps of Optimal Interpolation data assimilation updating surface velocity components with generated pseudo measurement by using Monte Carlo simulation is listed as follows:

- Run the free model from 01/10/2011 to 07/10/2011, calculate the background error variance, the error variance of measurement field is averaged on space using obtained background error variance, then calculate the Kalman gain with equation (3)
- Generate noise following normal distribution N (0, 16) and N (0, 9) for producing pseudo measurement, because surface velocity component are calculated separately in our numerical model, normal distribution noise N (0, 16) is for u component of surface velocity, normal distribution noise N (0, 9) is for v component of surface velocity.
- Control run is undertaken from 14/10/2011 to 20/10/2011 with no data assimilation, the time interval of its surface velocity components output in data assimilation domain is sixty minutes
- Add the generated noise to corresponding velocity components from Control run by using Monte Carlo simulation, the new dataset is the pseudo measurement field, then interpolate the pseudo measurement on time to every five minutes
- Update the surface velocity states every five minutes with the generated pseudo measurement by utilising Optimal Interpolation data assimilation scheme in the square data assimilation domain

Since the model takes about three days to be stable, the output of surface velocity components from the free run and data assimilation process are only taken during the last four simulation days.

4.1 BACKGROUND ERROR COVARIANCE (FORECAST ERROR COVARIANCE)

In Optimal Interpolation data assimilation scheme, the background error covariance is static, which means that the background error covariance is a constant matrix. It is calculated before the data assimilation process is performed (Robert, Blayo and Verron, 2005). According to the statistical relationship between the covariance and correlation coefficient, the background error covariance

could be expressed as (Oke, 2002):

$$\text{cov}(e_i, e_j) = \rho_{ij} \times \sigma_i \times \sigma_j , \tag{4}$$

where e_i, e_j are the errors at different locations, σ_i, σ_j are the standard deviation of errors at different locations, ρ_{ij} is the spatial correlation function, which is defined based on Gaussian function $\rho_{ij} = \alpha \exp(-\frac{d^2}{L^2})$, α is the scale factor, L is the correlation length, d is the spatial distance between two points. Our interest is to update the surface velocity components u and v. The appropriate parameters α and L are determined when the Root Mean Square Error reaches minimum.

According to equation (4), the background error covariance can be simplified as:

$$P^b = D^{\frac{1}{2}} C D^{\frac{1}{2}} , \tag{5}$$

D is the diagonal background state variance matrix describing the modelling error structure. The background state error e^b is computed from the difference over a short time interval ten minutes $(\Delta t = 10 \min s)$.

$$D = \text{var}(e^b) = E[e^b e^{bT}] = E[(e^b - \overline{e^b})(e^b - \overline{e^b})^T], \tag{6}$$

$$e^b = x_i^b - \overline{x_{\Delta t}^b} . \tag{7}$$

The overbar means the expected value.

e^b is the model error at different locations.

x_i^b is the background state at time step i.

C is the spatial correlation matrix. Every element is calculated from equation (4).

Since pseudo measurements are yielded from Control run by adding noise with Monte Carlo simulation, the pseudo measurement field has the same structure as background states of numerical modelling, the measurement operator H in equations (1) and (3) is an identity matrix.

In order to clearly show the improvement of sequential data assimilation schemes, in following Optimal Interpolation data assimilation, five minutes is chosen as the data assimilation interval.

4.2 ADJUSTMENT OF α AND L FOR OPTIMAL INTERPOLATION

For Optimal Interpolation data assimilation process, parameters α and L directly decide its effectiveness. Gu ((Gu, Woo and Kim, 2011)) had used different optimized values of the two parameters at different location to assimilate vertical current data to the unstructured grid ocean numerical model, the minimum and maximum of α was 1 and 4, the minimum and maximum value of

correlation length L was 20 km and 100 km. Ragnoli (Ragnoli, Zhuk, Donncha, Suits and Hartnett, 2012) assimilated the High Frequency radar surface current data of Galway bay in numerical model, the scaling factor war chosen as 100 and correlation length was 0.3 km, their data assimilation area was the whole inner Galway bay. For different research area and using different types of measurement, the optimal parameters of scaling factor and correlation length are different. In our research, a variety of test cases with different values of these parameters are investigated, RMSE is employed to measure the degree of their match. For these tests, assimilation interval is five minutes. Firstly, the RMSE is calculated on space (961 grids) every five minutes, then it is averaged on time (see Table 1 and Table 2).

TABLE 1 RMSE of u component

a/L	0.15 km	0.30 km	0.45 km	0.75 km	1.05 km	1.50 km	3.00 km
1	0.0523	0.0508	0.0508	0.0538	0.0605	0.0769	0.1910
2	0.0550	0.0524	0.0535	0.0642	0.0846	0.1292	0.3770
3	0.0570	0.0541	0.0573	0.0754	0.1094	0.1834	0.5705

TABLE 2 RMSE of v component

a/L	0.15 km	0.30 km	0.45 km	0.75 km	1.05 km	1.50 km	3.00 km
1	0.0383	0.0373	0.0372	0.0393	0.0438	0.0556	0.1947
2	0.0404	0.0385	0.0392	0.0472	0.0630	0.0994	0.4089
3	0.0418	0.0398	0.0418	0.0561	0.0833	0.1466	0.6074

From the RMSE (u) and RMSE (v) in these cases, when $\alpha = 1.0$ and $L = 0.45km$, both of them are minimum, so these values are employed in our Optimal Interpolation data assimilation.

5 Results

The goal of Optimal Interpolation data assimilation is to enhance the modelling prediction capability referring to the measurement trajectory. Surface current maps at certain time steps are displayed and statistical comparison of surface velocity components time series for reference points is shown in Taylor diagrams. The surface current maps of assimilation model field are compared with control model field with no data assimilation process and pseudo measurement field at certain time steps. Here, only the data assimilation domain is displayed.

In Figure 2-4, the left panels show surface current map at t=4.0 days, the right panels show surface current map at t=6.0 days. For both surface current map at t=4.0 days and t=6.0 days, compared with results from original model in Figure 3, Figure 4 shows that the Optimal Interpolation data assimilation process absorbs useful information from pseudo measurement into numerical model, since the consideration of observation error in Optimal Interpolation data assimilation process, the pseudo measurement is not fully projected into the numerical model, the assimilation model just assimilated the basic trend of pseudo measurement into assimilation model. The pseudo measurement is generated by adding normal distribution noise into the results from Control model, the general trend in the data assimilation domain is chaotic. We use this way to test the sensitivity reflection of our Optimal Interpolation data assimilation process when this kind of pseudo measurement is used for update. The reason behind this is that observation data in real world is always noisy. Generally, Optimal Interpolation works well when chaotic pseudo measurement is used for assimilation, and compared with Control model, general trend of assimilation model in data assimilation domain is closer to pseudo measurement trajectory.
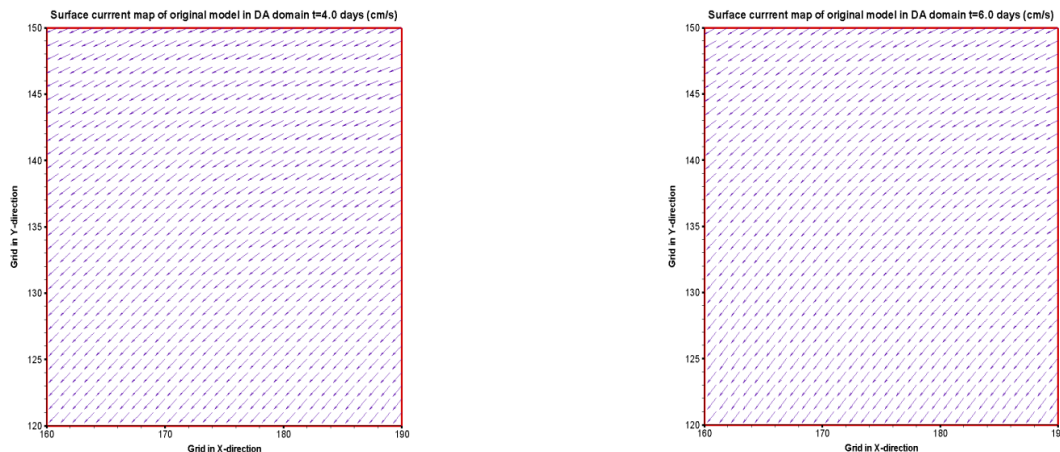
FIGURE 2 Surface current map with no data assimilation from Control model (t=4.0 days and t=6.0 days)
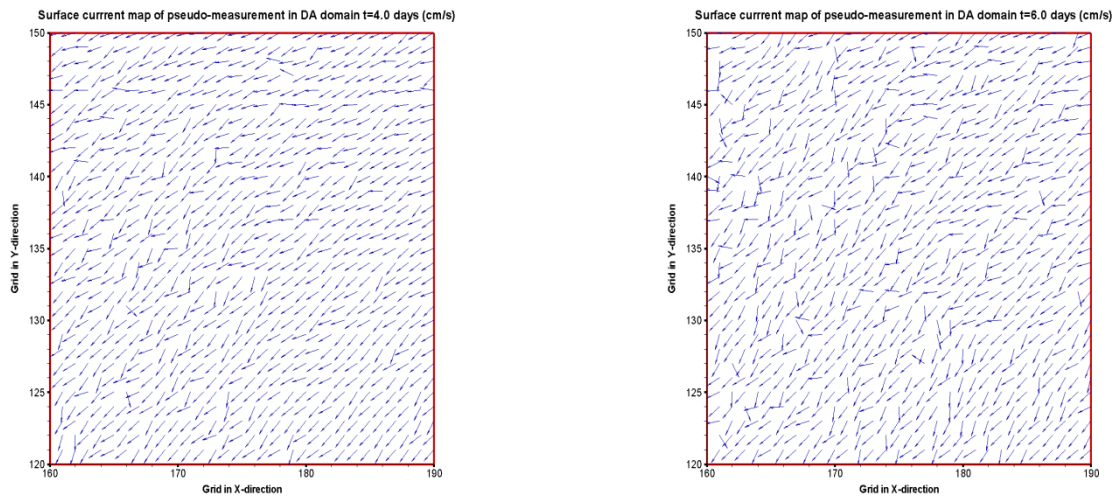
FIGURE 3 Surface current map of pseudo measurement (t=4.0 days and t=6.0 days)
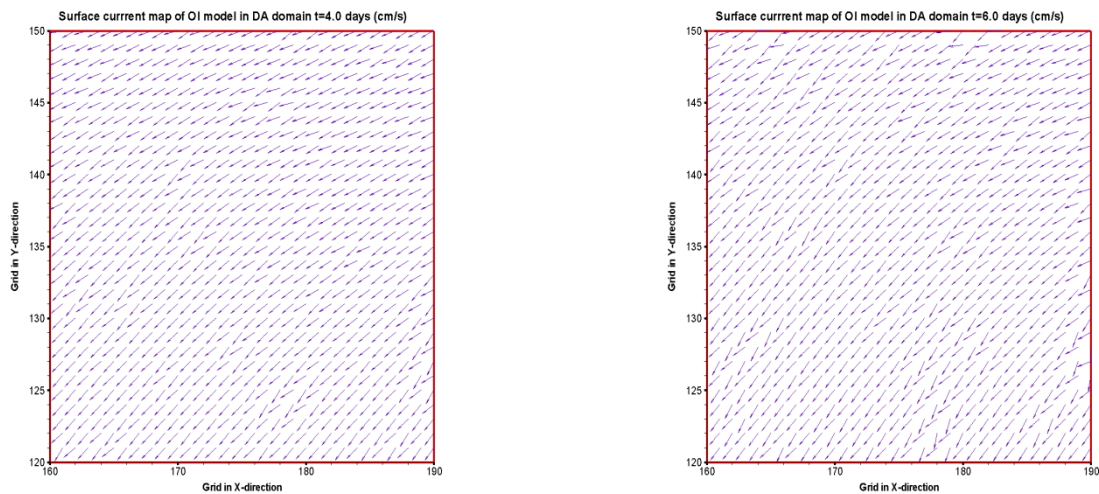
FIGURE 4 Surface current map with Optimal Interpolation data assimilation (t=4.0 days and t=6.0 days)

In order to further assess the effectiveness of Optimal Interpolation data assimilation, Taylor diagram (see Figure 5-12) of two inside reference point C, point D and two boundary reference (point A and point B) are displayed (see Figure 1). Data of these figures is from t=3.0 days to t=7.0 days. Taylor diagram is a graphical way to summarise degree of match between observation and reference models, their statistic (correlation, centred root-mean-square difference and standard deviation) could be concisely shown in terms of each model's position in the diagram (Taylor, 2001). In the following Taylor diagrams, the blue point means the results of data assimilation model, the red point stands for the results of control model with no data assimilation process from Control run, the hollow black dot on the horizontal axis is the measurement state. The centred root-mean-square difference between the modeming results and observed patterns is proportional to the distance to the point on the x-axis identified as measurement. The dotted contours indicate the RMS values. The dotted line from the origin to arch shows the correlation relationship between observation and modelling states. The standard deviation of the modelling results is proportional to the radial distance from the origin. Generally, the RMS values of assimilation states (blue point) is smaller than Control modelling results (red line), especially for point B, which means that the Optimal Interpolation data assimilation process makes the numerical modelling takes useful measurement information into the dynamic system. There is not obvious improvement of correlation relationship for point *A* and point *D*, but for point *B* and *v* component of point *C*, in other words, the assimilation process renders the model have a closer correlation relationship with measurement states. For the standard deviation, since the pseudo measurement is produced by adding normal distribution noise to the results from original modelling. Surface velocity components of four points time series are outputted at the exact data assimilation time step with five minutes assimilation interval. The chaotic pseudo measurement is not smooth comparing with results of the Control run, so the assimilation model could not show smaller standard deviation when noisy pseudo measurement is used. This also proves that the quality of measurement is of great importance for our

data assimilation, although the measurement error is considered in the Kalman gain, check of measurement data is needed. From the below Taylor diagrams, the majority of the Taylor diagrams show that the Optimal Interpolation data assimilation process improves the modelling prediction ability, making the model states closer to observation states, which means that Optimal Interpolation data assimilation is an effective tool to enhance the numerical modelling by blending the available measurement data.

Since the surface u and v pseudo measurement are assimilated in the numerical modelling respectively, the impact of data assimilation on the direction of total velocity is also investigated. Taylor diagrams of these reference points direction (t=3.0~7.0 days) are shown as follows. Except for point A, the correlation relationship between measurement states and assimilation states is enhanced and the RMS values are reduced with assimilation.
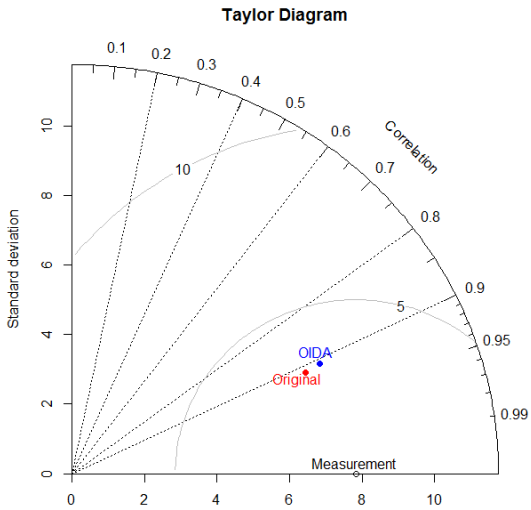
FIGURE 5 Taylor diagram of u component at point A
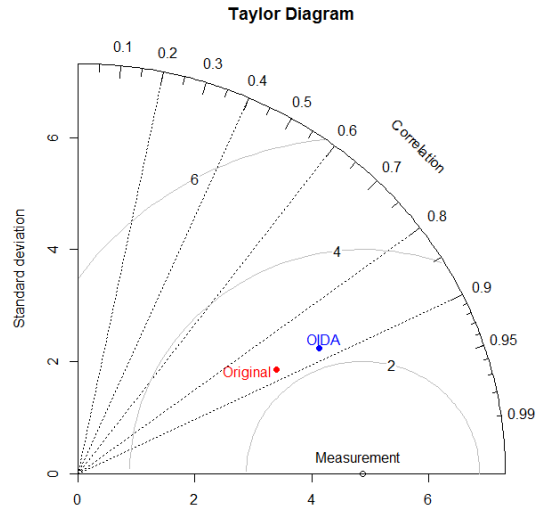
FIGURE 6 Taylor diagram of v component at point A
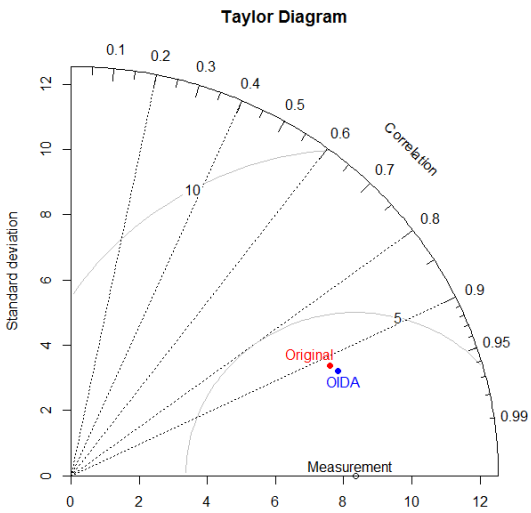
FIGURE 7 Taylor diagram of u component at point B
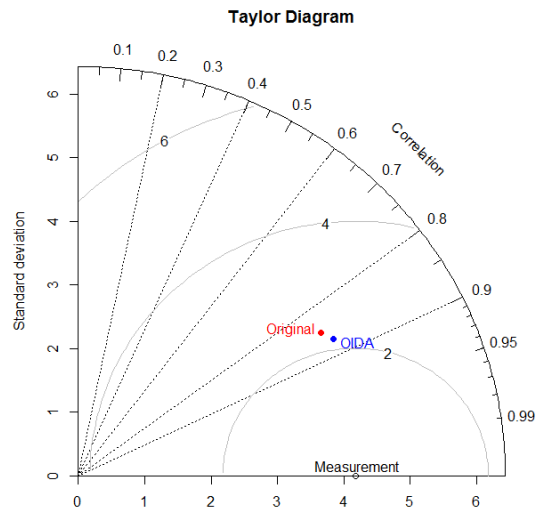
FIGURE 8 Taylor diagram of v component at point B
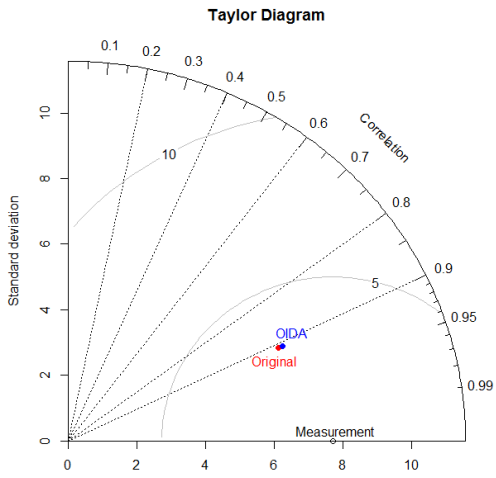
FIGURE 9 Taylor diagram of u component at point C
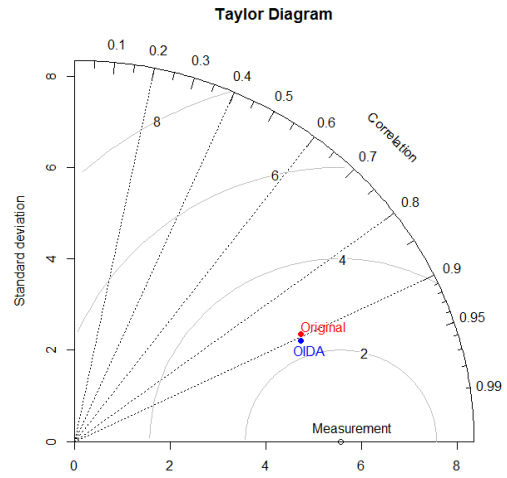
FIGURE 10 Taylor diagram of v component at point C
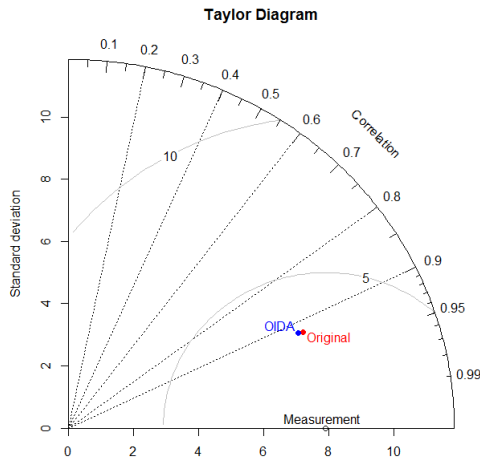
FIGURE 11 Taylor diagram of u component at point D
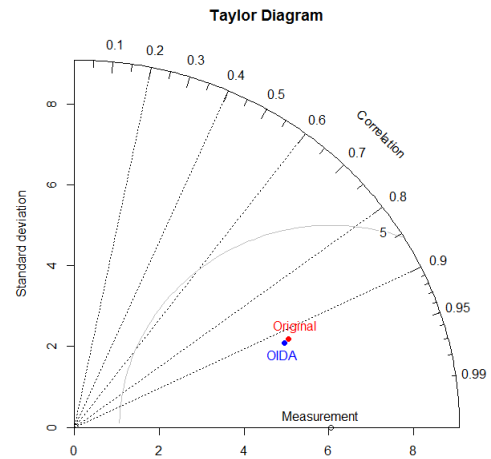
FIGURE 12 Taylor diagram of v component at point D
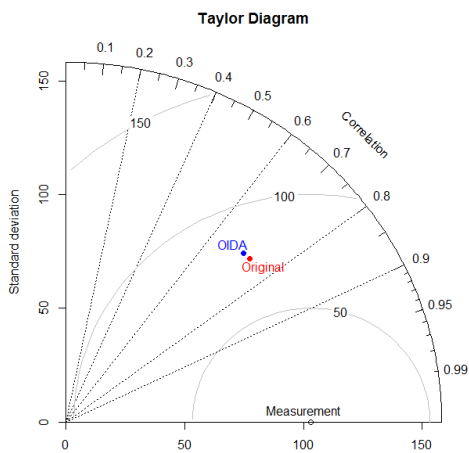
FIGURE 13 Direction Taylor diagram at point A
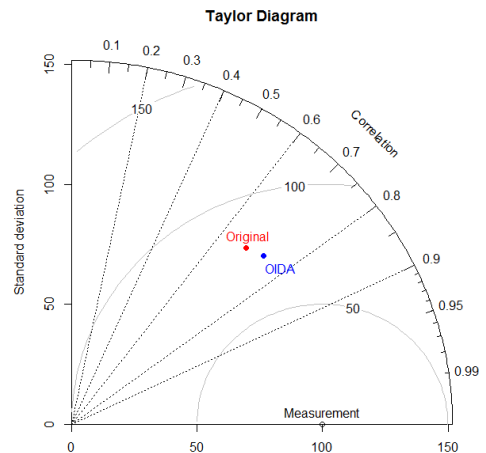
FIGURE 14 Direction Taylor diagram at point B
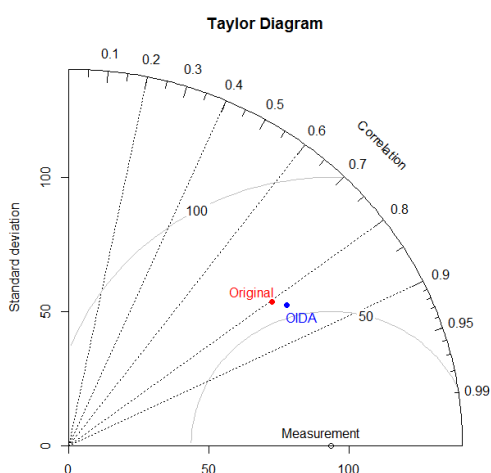
FIGURE 15 Direction Taylor diagram at point C
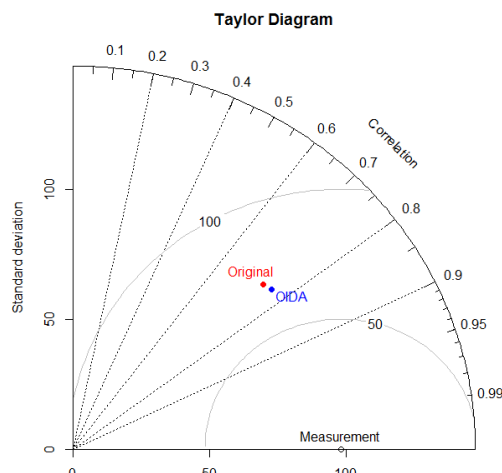
FIGURE 16 Direction Taylor diagram at point D

6 Conclusions

Taylor diagram at reference points and surface current maps in data assimilation domain showed that application of Optimal Interpolation to update model background states with pseudo measurement improves the modelling prediction ability in data assimilation domain, which means the method to calculate the background error covariance in our data assimilation system is meaningful. The improvement of Optimal Interpolation data assimilation is not obvious or the added analysis increment contaminates the background states in certain area or at few points (point A, v component). This is due to the modelling error covariance could not well stand for the development the modelling error, namely the background error covariance is stationary (Oke, Brassington, Griffin and Schiller, 2010, Counillon and Bertino, 2009, Oke, 2002). For further research, authors are trying to develop an operational real time forecasting surface current data assimilation system, real in situ measurement data will be used to update the background state in the following work.

The background error covariance was calculated from a free run, the model error was defined by subtracting the mean of background states over ten minutes from the model states. The time series improvement of surface velocity components at four inside reference points during the last four simulation days is displayed in Taylor diagram, the shown statistical values in Taylor diagrams depict that the model states are closer to the measurement trajectory when the Optimal Interpolation data assimilation is applied. The surface current map at certain time steps describe the Optimal Interpolation data assimilation process has assimilated the useful information from measurement into the model. When comparing with the original model that contained no data assimilation process, the general regional tends to follow the measurement trend after assimilation, which proves that the method used to compute the background error covariance is reasonable and Optimal Interpolation data assimilation works when pseudo measurement is used to update the model states. Optimal Interpolation data assimilation scheme of updating with pseudo measurement does improve the model prediction ability, which proves that this new way of computing background error covariance is efficient.

Acknowledgements

References

[1] Robinson A R, Lermusiaux P F J 2000 Overview of data assimilation Harvard Reports in Physical/Interdisciplinary (Ocean Sciences); The Division of Engineering and Applied Sciences, Harvard University: Cambridge, Massachusetts (NUMBER 62).
[2] Moore A M, Arango H G, Broquet G, Powell B S, Weaver A T, Zavala-Garay J 2011 The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems *Progress In Oceanography* **91**(1) 34-49
[3] Ma Z, Zheng Y, Zhong S, Zou X 2014 Study of maneuvering target tracking algorithm based on Kalman filter and ANFIS *Computer Modelling & New Technologies* **18**(1) 31-7
[4] Zaron E D 2009 *Introduction to Ocean Data Assimilation*
[5] Dong J, Xue M 2012 Assimilation of radial velocity and reflectivity data from coastal WSR-88D radars using an ensemble Kalman

filter for the analysis and forecast of landfalling hurricane Ike *Q J Roy Meteor Soc* **139**(671) 467-87

[6] Gu B, Woo S, Kim S 2011 Application of Optimal Interpolation to Vertical Current Data in the Ieo Island of Korea Using Unstructured Grid Ocean Numerical Model (FVCOM) *OCEANS* 2012 - YeosuYeosu, 1-8

[7] Rienecker M M 1991 Ocean Data Assimilation Using Optimal Interpolation With a Quasi-Geostrophic Model *Journal of Geophysical Research* **96** 93-103

[8] Oke P R, Brassington G B, Griffin D A, Schiller A 2010 Ocean data assimilation a case for ensemble optimal interpolation *Australian Meteorological and Oceanographic Journal* **59** 67-76

[9] Counillon F, Bertino L 2009 Ensemble Optimal Interpolation: multivariate propertiies in the Gulf of Mexico *Tellus A* **61**(2) 296-308

[10] Oke P R, Brassington G B, Griffin D A, Schiller A 2008 The Bluelink ocean data assimilation system (BODAS) *Ocean Modelling* **21**(1-2) 46-70

[11] Srinivasan A, Chassignet E P, Bertino L, Brankart J M, Brasseur P, Chin T M, Counillon F, Cummings J A, Mariano A J, Smedstad O M, Thacker W C 2011 A comparison of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM): Twin experiments with static forecast error covariances *Ocean Modelling* **37** 85-111

[12] Carton J A, Giese B S 2008 A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA) *Monthly Weather Review* **136**(8) 2999-3017

[13] Doucet A, de Freitas N, Gordon N 2001 *Sequential Monte Carlo methods in practice* Springer

[14] Kalnay E 2002 *Atmospheric Modeling, Data Assimilation and Predictability* Cambridge University Press.

[15] Robert C l, Blayo E, Verron J 2006 Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean *Proc., Modelling, observing, and forecasting sea level, ocean tides and ocean circulation; reviews and recent progress Special issue in honour and in memory of Christian Le Provost, Springer* 624-33

[16] Oke P R 2002 Assimilation of surface velocity data into a primitive equation coastal ocean model *Journal of Geophysical Research* **107**(C9)

[17] Ragnoli E, Zhuk S, Donncha F O, Suits F, Hartnett M 2012 An Optimal Interpolation Scheme for Assimilation of HF Radar Current Data into a Numerical Ocean Model *Hampton Road* **VA** 1 - 5

[18] Taylor K E 2001 Summarizing multiple aspects of model performance in a single diagram *Journal of Geophysical Research* **106**(D7) 7183.

Authors

Lei Ren, born on June 10, 1985, Hubei province, China

Current position, grades: PhD student
University studies: Master degree was earned in major of harbour, coastal and offshore engineering, Hohai University in 2011. He began his PhD research from October, 2011 in National University of Ireland, Galway
Scientific interest: Hydrodynamic modelling, data assimilation

Stephen Nash, born on June 4, 1976, Ireland

Current position, grades: lecturer
University studies: Ph.D. degree was earned in major of civil engineering, National University of Ireland, Galway in 2010
Scientific interest: physical and computer modelling of oceanic, estuarine and freshwater systems

Michael Hartnett, born on August 21, 1959, Birmingham, England

Current position, grades: professor
University studies: Ph.D. degree was earned in major of civil engineering, Trinity College Dublin in 2000.
Scientific interest: marine modelling: tidal energy; water quality; climate change; coastal flooding
Publications: 39 journal publications and 61 conference publications

Logistics outsourcing and selecting of logistics service provider of the e-commerce companies: a fuzzy TOPSIS approach

Yanwen Wang*, Xiuju Gao, Liming Yang

College of Finance, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P. R. China

Abstract

Logistics outsourcing has become the development trend of enterprise logistics operations. A good logistics service provider can improve the customer satisfaction, as also as reducing the cost of the whole supply chain, so it is very important to evaluate the logistics outsourcing service for the corn companies in the supply chain. This paper is an attempt to identify the main factors of selecting satisfactory logistics service provider of the electronic commerce (e-commerce) companies in China. GRA and Fuzzy TOPSIS are employed to evaluate the service of the logistics companies. The managerial implications are also discussed in the last section.

Keywords: logistics outsourcing, e-commerce, Fuzzy TOPSIS, evaluation

1 Introduction

Supply chain management involves the design and management of seamless, value-added processes across organizational boundaries to meet the real needs of the customers [1-3]. Logistics outsourcing is rapidly growing through the whole world. More than 70% of companies in Western Europe, USA and Asia Pacific have outsourcing experience in a pattern expanding from basic transportation to full logistics network control [4, 5].

In the early 1980s, logistics services in the outsourcing market were confined to the traditional activities such as transportation and warehousing. In the 1990s, a number of network players began to provide a wider geographic coverage of their transport networks, and many value adding activities such as labelling and sorting were introduced. In the late 1990s, a number of players from areas as information technology, management consultancy and financial services began working together with logistics service providers. This period saw the creation of a new service, the 'supply chain solution', also called 'fourth-party logistics (4PL)', where a logistics service provider (LSP) is hired to manage a customer's complete logistics network [6-8].

As the development of the electronic-commerce (e-commerce), an increasingly people prefer online shopping because of the low price and convenience. Most of the logistics outsourcing performance-related studies focused on the cost saving [4, 9, 10] and traditional logistics service [11, 12], few empirical studies have reported on the selection of the logistics service provider of the e-commerce companies, which is different from the traditional logistics service in terms of the payment technology, which is the primary motivation of this research.

This paper is organized as follows. The next section introduces the related literature about logistics service provider selection. Following is a brief introduction about the fuzzy TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method used in this research. Section 4 describes an empirical analysis of logistics service provider selection of an e-commerce company Z. Finally, major issues and challenges for logistics outsourcing are identified and discussed along with the related managerial implications.

2 Literature review

There are many researches focused on the performance of the logistics service providers. Kasilingamr [13] argue that four factors for the third party logistics service supplier to choose: the perceived performance of logistics suppliers, the perception ability, the price, the strategy and external environment using the factor analysis method. Ma etc. [14] used fuzzy analytic hierarchy process (F-AHP) to choose partners. Yahya and Kingsman [15] suggest evaluation index system including quality, response delivery and performance of financial management technical ability and facilities through the investigation and AHP. Guo [16] establish an AHP judgment matrix of supplier evaluation with quality, price, technical ability and distribution reliability. Jiang and Han [17] set up evaluation index system including quality, price, delivery, service, product development and production, external environment, and other (sales and marketing staff in general) on the comprehensive analysis of the service. Ma [11] proposed 9 indexes on supplier selection standards: the product quality, the price, the post-sale service, the technical level, the geographical position, supply capacity, economic benefit, delivery and market effect. Ma et al. [11] chose

---

* *Corresponding author* e-mail: 125721815@qq.com

Operation Research and Decision Making

three common indexes, such as quality, cost and delivery time, and set up a weight correlation analysis model of supplier selection and evaluation. Zhong et al. [18] suggest that the specific vendor selection indexes should consider four aspects such as technical level, management ability, and service level and management environment to make the supply chain performance maximization according to the design principle. Zhou et al. [12] set up a customer satisfaction index system from the customer's point of view to evaluate the third party logistics enterprise. It is proved that the system has the high homogeneity, the reliability and validity of the structures. Sun [19] construct a third-party logistics operation efficiency evaluation index system including four aspects, such as the input-output efficiency, the equipment utilization efficiency, quality assurance, efficiency, market competition efficiency. Hsiao et al. [20] identify and analyse the outsourcing of four levels of logistics activities: transportation (level 1), packaging (level 2), transportation management (level 3), and distribution network management (level 4). The evaluation index system of Peng [21] including logistics cost, the logistics operation efficiency, the basic qualities of service suppliers and logistics technology level has more targeted and practicability.

The TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method selected for the data analysis in this research was first proposed in 1981 [22]. Many proposed numerical examples have shown that TOPSIS can avoid some weaknesses of the existing multi-attribute methods [23]. Fuzzy TOPSIS is widely employed to solve the multi-criteria evaluation problems [24-26].

In summary, the evaluation of logistics service providers both in theory and in practice has proven to be very important and quite complex, and there have been limited researches in the current literature focusing on the selection of the logistics service provider of the e-commerce companies.

3 Methodology

3.1 FUZZY SETS AND FUZZY NUMBERS

First, it is necessary to review the related Fuzzy Theory:
Definition 1: A Fuzzy set $\tilde{a}$ in a universe of discourse X is characterized by a membership function $\mu_{\tilde{a}}(x)$ which associates with each element x in X, a real number in the interval [0,1]. The function of $\mu_{\tilde{a}}(x)$ is termed the grade of membership of x in $\tilde{a}$ [27]. The present study uses triangular Fuzzy numbers. $\tilde{a}$ can be defined by a triplet (a_1, a_2, a_3). Its conceptual schema and mathematical form are shown as below:

$$\mu_{\tilde{a}}(x) \begin{cases} 0 & x \le a \\ \dfrac{x - a_1}{a_2 - a_1} & a_1 < x \le a_2 \\ \dfrac{a_3 - x}{a_3 - a_2} & a_2 < x \le a_3 \\ 1 & x > a_3 \end{cases}.$$

Definition 2: Let $\tilde{a} = (a_1, a_2, a_3)$ and $\tilde{b} = (b_1, b_2, b_3)$ be two triangular fuzzy numbers. According to Wang [26], a distance measure function $(\tilde{a}, \tilde{b})$ can be defined as below:

$$d(\tilde{a}, \tilde{b}) = \sqrt{\frac{1}{3}[(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2]}.$$

Definition 3: Let a triangular Fuzzy number $\tilde{a}$, then α-cut defined as below:

$$\tilde{a}_\alpha = [(a_2 - a_1)\alpha + a_1, a_3 - (a_3 - a_2)\alpha].$$

Definition 4: Let $\tilde{a} = (a_1, a_2, a_3)$, $\tilde{b} = (b_1, b_2, b_3)$ be two triangular Fuzzy number and $\tilde{a}_\alpha$, $\tilde{b}_\alpha$ be α-cut, $\tilde{a}$ and $\tilde{b}$, then the method is defined to calculate the divided between $\tilde{a}$ and $\tilde{b}$ as follows (Kwang, 2005):

$$\frac{\tilde{a}_\alpha}{\tilde{b}_\alpha} = \left[\frac{(a_2 - a_1)\alpha + a_1}{-(b_3 - b_2)\alpha + b_3}, \frac{-(a_3 - a_2)\alpha + a_3}{(b_2 - b_1)\alpha + b_1} \right].$$

When $\alpha = 0$:

$$\frac{\tilde{a}_0}{\tilde{b}_0} = \left[\frac{a_1}{b_3}, \frac{a_3}{b_1} \right].$$

When $\alpha = 1$:

$$\frac{\tilde{a}_1}{\tilde{b}_1} = \left[\frac{(a_2 - a_1) + a_1}{-(b_3 - b_2) + b_3}, \frac{-(a_3 - a_2) + a_3}{(b_2 - b_1) + b_1} \right],$$

$$\frac{\tilde{a}_1}{\tilde{b}_1} = \left[\frac{a_2}{b_2}, \frac{a_2}{b_2} \right].$$

So the approximated value of $\tilde{a} / \tilde{b}$ will be:

$$\frac{\tilde{a}}{\tilde{b}} = \left[\frac{a_1}{b_3}, \frac{a_2}{b_2}, \frac{a_3}{b_1} \right].$$

Definition 5: Assuming that both $\tilde{a} = (a_1, a_2, a_3)$ and $\tilde{b} = (b_1, b_2, b_3)$ are real numbers, the distance measurement $d(\tilde{a}, \tilde{b})$ is identical to the Euclidean distance.

The basic operations on Fuzzy triangular numbers are as follows [28]:

For approximation of multiplication:

$$\tilde{a} \otimes \tilde{b} = (a_1 \times b_1, a_2 \times b_2, a_3 \times b_3).$$

For addition:

$$\tilde{a} \oplus \tilde{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3) .$$

3.2 GRA (GREY RELATIONAL ANALYSIS)

Below is a briefly review of relevant definitions and the calculation procedure for the GRA approach.

GRA uses several small sub-problems to present the decision problem, and the problem is decomposed into a hierarchy with a goal at the top, criteria and sub-criteria at levels and sub-levels and decision alternatives at the bottom of the hierarchy.

The comparison matrix involves the comparison in pairs of the elements of constructed hierarchy. The aim is to set their relative priorities with respect to each of the elements at the next higher level.

$$D = \begin{matrix} & C_1 & C_2 & C_3 & ... & C_n \\ C_1 & \begin{bmatrix} x_{11} & x_{12} & x_{13} & ... & x_{1n} \\ C_2 & x_{21} & x_{22} & x_{13} & ... & x_{2n} \\ C_3 & x_{31} & x_{32} & x_{33} & ... & x_{3n} \\ ... & ... & ... & ... & ... & ... \\ C_m & x_{m1} & x_{m2} & x_{m3} & ... & x_{mn} \end{bmatrix} \end{matrix},$$

where x_{ij} is the degree preference of i-th criterion over j-th criterion. Before the calculation of vector of priorities, the comparison matrix has to be normalized into the range of [0,1] by the equation below:

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} .$$

Step 1: The consistency ratio need to be identified to reflect the consistency of the decision maker's judgments during the evaluation phase.

$$CI = \frac{\lambda_{max} - N}{N - 1} ,$$

where CI is the consistency ratio, λ_{max} is the principal eigenvalue of the judgement matrix and N is the order of the judgement matrix. The consistency ratio should be lower than 0.10 to accept the AHP results as consistent.

Step 2: In the next step, transform r_{ij} into the fuzzy numbers.

Step 3: Calculate the average of the elements of each rows from matrix obtained from step 4, by Definition 4.

3.3 FUZZY MEMBERSHIP FUNCTION

In the evaluating process, the weights expressed with the linguistic terms, represent the important degrees of criteria from experts via surveys on subjective assessments. These linguistic terms are categorized into very low (VL), low

(L), medium (M), high (H) and very high (VH). Assume that all linguistic terms can be transferred into triangular fuzzy numbers, and these fuzzy numbers are limited in [0, 1]. As a rule of thumb, each rank is assigned an evenly spread membership function that has an interval of 0.30 or 0.25.

Based on assumptions above, a transformation table can be found as shown in Table 1. Figure 1 illustrates the Fuzzy membership function [28].

TABLE 1 Transformation for Fuzzy Membership Functions

Rank function	Sub-criteria grade	Membership
Very Low (VL)	1	(0.00,0.10,0.25)
Low (L)	2	(0.15,0.30,0.45)
Medium (M)	3	(0.35,0.50,0.65)
High (H)	4	(0.55,0.70,0.85)
Very High (VH)	5	(0.75,0.90,1.00)
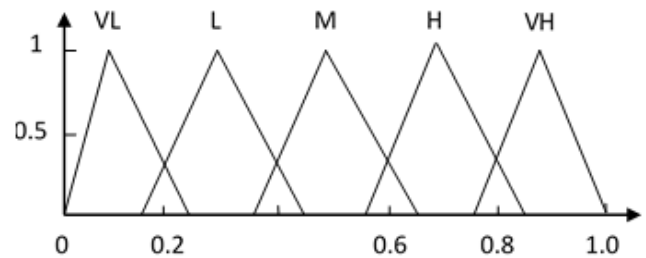
FIGURE 1 Fuzzy triangular membership functions

3.4 FUZZY TOPSIS MODEL

To describe the evaluation method clearly, the procedure of fuzzy TOPSIS is presented as below. It is formulated that a Fuzzy Multiple Criteria Decision Making (FMCDM) problem about the comparative evaluation of the websites of those laptop manufacturers. The FMCDM problem can be concisely expressed in matrix format as follows:

$$\begin{matrix} & C_1 & C_2 & C_3 & ... & C_n \\ A_1 & \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} & ... & \tilde{x}_{1n} \\ A_2 & \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{13} & ... & \tilde{x}_{2n} \\ A_3 & \tilde{x}_{31} & \tilde{x}_{32} & \tilde{x}_{33} & ... & \tilde{x}_{3n} \\ ... & ... & ... & ... & ... & ... \\ A_m & \tilde{x}_{m1} & \tilde{x}_{m2} & \tilde{x}_{m3} & ... & \tilde{x}_{mn} \end{bmatrix} \end{matrix}$$

$$\tilde{W} = [\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_n] ,$$

where x_{ij}, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$ and $\tilde{w}_j$, $j = 1, 2, ..., n$ are linguistic triangular Fuzzy numbers, $\tilde{x}_{ij} = (a_{ij}, b_{ij}, c_{ij})$ and $\tilde{w}_j = (a_{j1}, b_{j2}, c_{j3})$. The normalized Fuzzy decision matrix is denoted by $\tilde{R} = [\tilde{r}_{ij}]_{m \times n}$.

The weighted Fuzzy normalized decision matrix is shown as follows:

$$
V = \begin{bmatrix} \tilde{v}_{11} & \tilde{v}_{12} & \tilde{v}_{13} & ... & \tilde{v}_{1n} \\ \tilde{v}_{21} & \tilde{v}_{22} & \tilde{v}_{13} & ... & \tilde{v}_{2n} \\ \tilde{v}_{31} & \tilde{v}_{32} & \tilde{v}_{33} & ... & \tilde{v}_{3n} \\ ... & ... & ... & ... & ... \\ \tilde{v}_{n1} & \tilde{v}_{n2} & \tilde{v}_{n3} & ... & \tilde{v}_{nn} \end{bmatrix} =
$$

$$
\begin{bmatrix} \tilde{w}_1\tilde{r}_{11} & \tilde{w}_2\tilde{r}_{12} & \tilde{w}_3\tilde{r}_{13} & ... & \tilde{w}_n\tilde{r}_{1n} \\ \tilde{w}_1\tilde{r}_{21} & \tilde{w}_2\tilde{r}_{22} & \tilde{w}_3\tilde{r}_{13} & ... & \tilde{w}_n\tilde{r}_{2n} \\ \tilde{w}_1\tilde{r}_{31} & \tilde{w}_2\tilde{r}_{32} & \tilde{w}_3\tilde{r}_{33} & ... & \tilde{w}_n\tilde{r}_{3n} \\ ... & ... & ... & ... & ... \\ \tilde{w}_1\tilde{r}_{m1} & \tilde{w}_2\tilde{r}_{m2} & \tilde{w}_3\tilde{r}_{m3} & ... & \tilde{w}_n\tilde{r}_{mn} \end{bmatrix}.
$$

Given the above Fuzzy theory, the proposed Fuzzy TOPSIS procedure is then defined as follows:

Step 1:

Choose the $x_{ij}, i = 1, 2, ..., m; j = 1, 2, ..., n$ for alternatives with respect to criteria and $\tilde{w}_j, j = 1, 2, ..., n$ for the weight of the criteria.

Step 2: Construct the weighted normalized Fuzzy decision matrix V.

Step 3: Identify positive ideal (A^+) and negative ideal (A^-) solutions:

$$
A^+ = \left\{ \tilde{v}_1^+, \tilde{v}_2^+, ..., \tilde{v}_n^+ \right\}
$$
$$
= \left\{ (\max_i \tilde{v}_{ij} | i = 1, 2, ..., m), j = 1, 2, ..., n \right\},
$$

$$
A^- = \left\{ \tilde{v}_1^-, \tilde{v}_2^-, ..., \tilde{v}_n^- \right\}
$$
$$
= \left\{ (\min_i \tilde{v}_{ij} | i = 1, 2, ..., m), j = 1, 2, ..., n \right\}.
$$

Considering that the elements $\tilde{v}_{ij}$ are normalized positive triangular fuzzy numbers and their ranges belong to the closed interval [0,1], the positive ideal and negative ideal solutions can be defined as $\tilde{v}_j^+ = (1,1,1)$ and

$$
\tilde{v}_j^- = (0,0,0), j = 1, 2, ..., n. \tag{29}
$$

Step 4: Calculate separation measures. The distance of each alternative from A^+ and A^- can be identified as follows:

$$
d_i^+ = \frac{1}{n} \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^+), i = 1, 2, ..., m,
$$

$$
d_i^- = \frac{1}{n} \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^-), i = 1, 2, ..., m.
$$

Step 5: Calculate the similarities to ideal solution:

$$
CC_i = \frac{d_i^-}{d_i^+ + d_i^-}.
$$

Step 6: Rank preference order. Rank alternatives according to CC_i in descending order [28].

4 Solutions from GRA and TOPSIS analysis

The evaluation criteria in terms of logistics service providers of the e-commerce companies are show in Figure 2, and payment technology is added to the evaluation system identified by Peng [21], since this paper is focused on the evaluation of e-commerce companies' logistics service providers.
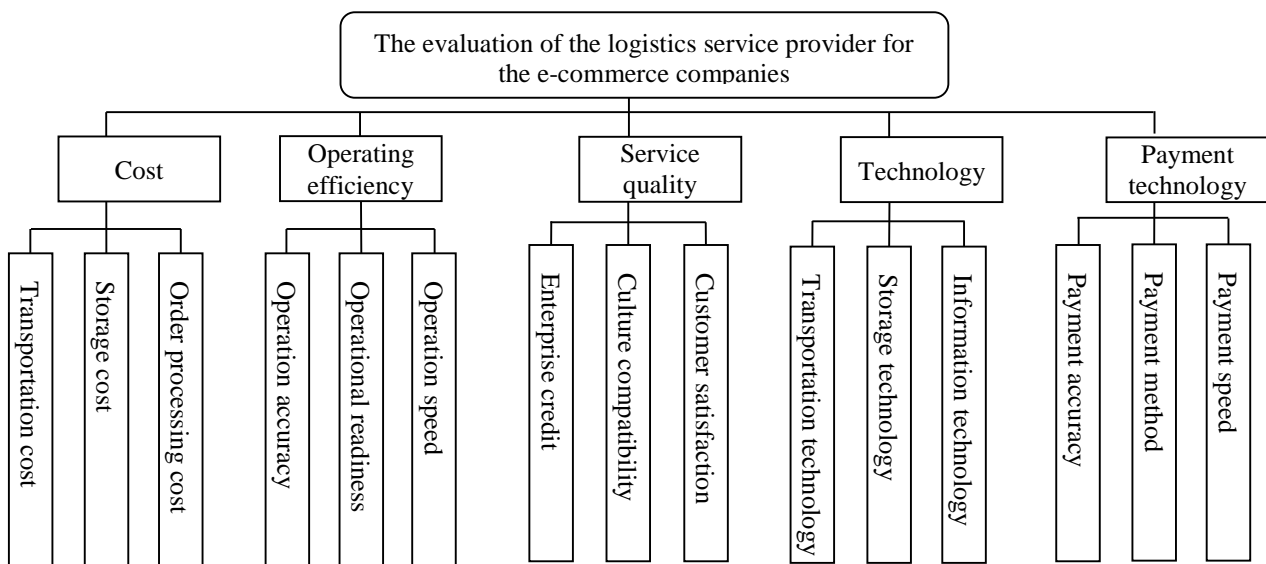
FIGURE 2 The evaluation of the logistics service provider for the e-commerce companies

Fuzzy TOPSIS, as a quantitative tool, is employed in this research. These specific measures are listed and named in Figure 2. The decision problem consists of three levels: at the highest level, the objective of the problem is situated while in the second level, the criteria are listed, and in the third level, the sub-criteria are listed.

$$R_{A1} = \begin{bmatrix} 1.0000 & 0.2904 & 0.2693 \\ 0.2894 & 1.0000 & 1.0000 \\ 0.2811 & 1.0000 & 1.0000 \end{bmatrix},$$

$$R_{A2} = \begin{bmatrix} 1.0000 & 0.5727 & 0.5727 \\ 0.5504 & 1.0000 & 1.0000 \\ 0.5727 & 1.0000 & 1.0000 \end{bmatrix},$$

$$R_{A3} = \begin{bmatrix} 1.0000 & 0.0990 & 1.0000 \\ 0.0990 & 1.0000 & 0.0921 \\ 1.0000 & 0.0986 & 1.0000 \end{bmatrix},$$

$$R_{A4} = \begin{bmatrix} 1.0000 & 0.2220 & 0.2008 \\ 0.2564 & 1.0000 & 0.5624 \\ 0.2559 & 0.5667 & 1.0000 \end{bmatrix},$$

$$R_{A5} = \begin{bmatrix} 1.0000 & 0.1741 & 0.3672 \\ 0.1925 & 1.0000 & 0.1541 \\ 0.4012 & 0.1650 & 1.0000 \end{bmatrix}.$$

According to the above matrices and the definition described earlier, the measures can be grouped into several clusters on the three aspects by threshold value $r = 0.60$. The final criteria are shown in Table 2. The weights of the criteria are driven from five experts as shown in Table 3. In the next step, the rates are asked to evaluate the four logistics outsourcing companies, and considering the limit of the page, part of the results are shown in Table 4.

TABLE 2 The evaluation criteria of the logistics service provider for the e-commerce companies

Goal	Aspects	Criteria
	C_1 Cost	SC_1 Transportation cost SC_2 Storage cost
	C_2 Operating efficiency	SC_3 Operation accuracy SC_4 Operational readiness
The logistics service provider evaluation for the e-commerce companies	C_3 Service quality	SC_5 Enterprise credit SC_6 Culture compatibility
	C_4 Technology level	SC_7 Transportation technology SC_8 Storage technology SC_9 Information technology
	C_5 Payment technology	SC_{10} Payment accuracy SC_{11} Payment method SC_{12} Payment speed

TABLE 3 The linguistic weights given by five experts

	Sub- Criteria	E1	E2	E3	E4	E5
cost	SC_1	VH	VH	H	VH	H
	SC_2	H	M	M	H	M
Operating efficiency	SC_3	H	M	H	H	VH
	SC_4	H	M	L	M	H
Service quality	SC_5	H	H	VH	H	H
	SC_6	M	L	L	L	VL
Technology level	SC_7	M	M	L	M	L
	SC_8	L	M	H	M	L
	SC_9	H	M	H	M	H
Payment technology	SC_{10}	H	M	H	L	M
	SC_{11}	H	M	H	H	H
	SC_{12}	M	L	L	VL	L

TABLE 4 Part of the evaluation results using Fuzzy linguistic variables

	Cost		Operating efficiency		Service quality		Technology level		
	Transport cost	Storage cost	Operation accuracy	Operational readiness	Enterprise credit	Culture compatibility	Transport tech.	Storage tech.	Information tech.
A1	VH	M	H	M	M	L	M	M	L
A2	H	H	H	H	H	M	H	M	H
A3	VH	H	H	VH	H	M	H	H	H
A4	H	M	M	M	M	L	M	M	M

Then the normalized decision matrix is then derived from the original data as follows:

The larger, the better type [28]:

$$r_{ij} = \frac{\left[x_{ij} - \min\left\{ x_{ij} \right\} \right]}{\left[\max\left\{ x_{ij} \right\} - \min\left\{ x_{ij} \right\} \right]}.$$

The smaller, the better type:

$$r_{ij} = \frac{\left[\max\{x_{ij}\} - x_{ij} \right]}{\left[\max\{x_{ij}\} - \min\{x_{ij}\} \right]} .$$

The Fuzzy linguistic variable is then transformed into a Fuzzy triangular membership function as shown in Table 5, and then the resulting Fuzzy weighted decision

matrix can be derived based on Table 5 and the weights identified before. As discussed in 3.3, the positive ideal and negative ideal solutions can be defined as $\tilde{v}_j^+ = (1,1,1)$ and $\tilde{v}_j^- = (0,0,0), j = 1, 2, ..., n$. The distance of each alternative from A^+ and A^-, as well as the similarities to an ideal solution, is obtained in Table 6.

TABLE 5 Part of the Fuzzy decision matrix

	A1	**A2**	**A3**	**A4**
Transportation cost	(0.75,0.90,1.00)	(0.55,0.70,0.85)	(0.75,0.90,1.00)	(0.55,0.70,0.85)
Storage cost	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Operation accuracy	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Operational readiness	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.75,0.90,1.00)	(0.35,0.50,0.65)
Enterprise credit	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Culture compatibility	(0.15,0.30,0.45)	(0.35,0.50,0.65)	(0.35,0.50,0.65)	(0.15,0.30,0.45)
Transportation technology	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Storage technology	(0.35,0.50,0.65)	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Information technology	(0.15,0.30,0.45)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Payment accuracy	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.35,0.50,0.65)
Payment method	(0.00,0.10,0.25)	(0.55,0.70,0.85)	(0.55,0.70,0.85)	(0.15,0.30,0.45)
Payment speed	(0.35,0.50,0.65)	(0.35,0.50,0.65)	(0.55,0.70,0.85)	(0.35,0.50,0.65)

TABLE 6 The distance of each alternative from A^+ and A^-

No.	d_i^+	d_i^-	CC_i
A_1	0.129	0.024	0.157
A_2	0.034	0.119	0.780
A_3	0.000	0.153	1.000
A_{17}	0.132	0.021	0.139

In order to see the result more clearly, the resulting Fuzzy TOPSIS analysis is shown in Figure 3.
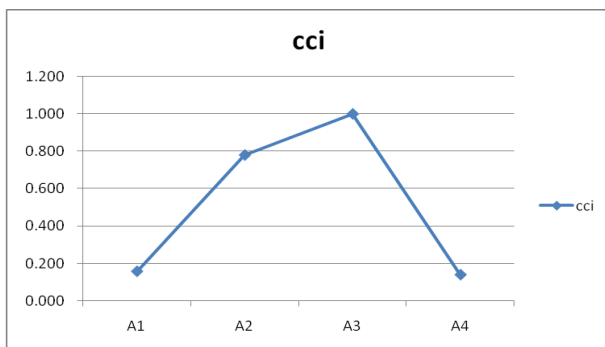
FIGURE 3 Summary of the evaluation of the four logistics outsourcing companies

It is can be seen from the Figure 3 that the third logistics outsourcing company is the best among the four companies to be identified as the logistics service provider.

5 Conclusions and suggestions for future research

This study is an attempt to identify the main factors of selecting satisfactory logistics service provider of the electronic commerce (e-commerce) companies in China.

GRA and Fuzzy TOPSIS are employed to evaluate the service of the logistics companies. According to the criteria weights derived from this section earlier, the relative top three important measures to evaluate a logistics outsourcing company are (1) SC_1 Transportation cost, and its weight = (0.67, 0.82, 0.94); (2) SC_5 Enterprise credit, and its weight = (0.59, 0.74, 0.88); and (3) SC_3 Operation accuracy, and its weight = (0.55,0.70,0.84). As such, when the corn company of the supply chain selects the logistics outsourcing companies, it should choose the logistics service company, which is low in transportation coast and has good enterprise credit and high operation accuracy.

Based on the results of this research, our recommendations for improving logistics service in terms of enhancing their effectiveness are: (1) selecting the high operation efficiency logistics company; (2) combing logistics service and payment service; and (3) strengthen the evaluation of the logistics outsourcing company.

Acknowledgments

References

[1] Christopher M L 1992 Logistics and Supply Chain Management *Pitman Publishing* London

[2] Ellram L M 1991 Supply chain management: the industrial organization perspective. International *Journal of Physical Distribution & Logistics Management* **21**(1) 13–22

[3] Fawcett S, Ellram L, Ogden J 2007. Supply Chain Management from Vision to Implementation *Pearson Education* New Jersey

[4] Capgemini 2007 Third-party logistics /http://www.at.cgey.com (accessed 11 October 2008)

[5] Lu Q, Meng F, Goh M 2014 Choice of Supply Chain Governance: Self-managing or Outsourcing? *International Journal of Production Economics* **154** 32-8

[6] Carbone V, Stone M 2005 Growth and relational strategies used by the European logistics service providers: rationale and outcomes *Transportation Research Part E* **41** 495–510

[7] Hertz S, Alfredsson M 2003 Strategic development of the third party logistics providers *Industrial Marketing Management* **32** 139–79.

[8] Lai K H 2004 Service capability and performance of logistics service providers *Transportation Research Part E* **40** 385–99

[9] Capgemini 2005 Third-party logisticsstudy /http://www.at.cgey.com (accessed 5 September2006)

[10] Lau K H, Zhang J 2006. Drivers and obstacles of out sourcing practices in China *International Journal of Physical Distribution & Logistics Management* **36**(10) 776–92

[11] Ma S, Wang X 2002 A study on the method of determining supplier evaluation index-weight *Industrial Engineering and Management* 5-8

[12] Zhou J, Wang L 2005 Analysis of influence factors of customer satisfaction on the third party logistics enterprise *Logistics Technology* 30-3

[13] Kasilingamr R 1998 Logistics and transportation *Kluwer Academic Publishers* 11-8

[14] Ma P, Zhu D, Ding Y 1999 Optimization selection member method based on fuzzy analytical hierarchy process (F-AHP) *Journal of Xi'an Jiaotong University* 108-10

[15] Yahya S, Kingsman B 1999 Vendor rating for an entrepreneur development programme: a case study using the analytic hierarchy process method *Journal of Operation Research Society* **50** 916930

[16] Guo Y, Li Y 1999 Analysis of strategic partnership friendship between manufacturer and supplier, Journal of Harbin University of Science and Technology 72-5

[17] Jiang H, Han W 2001 Selection standards and methods on supplier *Science-Technology and Management* **3** 69-70

[18] Zhong W Hou Q 2003 A study on evaluation indicator system and evaluation model of suppliers *The Journal of Quantitative & Technical Economics* 93-7

[19] Sun Y 2006 Constructing of the evaluation index system of third-party logistics operation efficiency *Special Zone Economy* 58-60

[20] Hsiao L, Kemp R G M, van der Vorst J G A J, Omta S W F 2010 A classification of logistic out sourcing levels and their impact on service performance: Evidence from the food processing industry *International Journal of Production Economics* **124** 75–86

[21] Peng J 2012 Selection of Logistics Outsourcing Service Suppliers Based on AHP *Energy Procedia* **17** 595-601

[22] Hwang C, Yoon K 1981 Multiple attribute decision making methods and applications *Springer-Verlag* Heidelberg 3-55

[23] Byun H, Lee K 2005 Decision support system for the selection of a rapid proto typing process using the modified TOPSIS method *International Journal of Advanced Manufacturing Technology* **26** 1338-47

[24] Muralidhar P, Ravindranath K, Srihari V 2013 The influence of GRA and TOPSIS for assortment of green supply chain management strategies in Cement industry *International Journal of Supply Chain Management* **2**(1) 49-55

[25] Zeki A, Rifat G O, 2012 Evaluating machine tool alternatives through modified TOPSIS and alpha-cut based fuzzy ANP *International Journal of Production Economics* **140**(2) 630-6

[26] Wang Z, Wang Y 2014 Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method Original Research Article *Expert Systems with Applications* **41**(6) 2824-31

[27] Zadeh L A 1965 Fuzzy sets *Information Control* **8** 338-53

[28] Yang T, Hung C 2007 Multiple-attribute decision-making methods for plant layout design problem *Robotics and Computer-integrated manufacturing* **23** 126-37

[29] Isiklar G, Buyukozkan G 2006. Using a multi-criteria decision making approach to evaluate mobile phone alternatives *Computer Standards and Interfaces* **29** 265-74

Authors

Yanwen Wang, born in July, 1977, Changli, Hebei, China

Current position, grades: lecturer at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei.
University studies: Master's degree of economics at University of International Business and Economics in 2010.
Scientific interest: economics, trade, education.
Publications: 1

Xiuju Gao, born in December, 1971, Changli, Hebei, China

Current position, grades: Associate Professor at Hebei Normal University of Science & Technology.
University studies: College of Finance.
Scientific interest: Master's degree of economics at Hebei University of Economics and Business, Shijiazhuang, Hebei in 2012.
Publications: 1.

Liming Yang, born in November, 1978, Changli, Hebei, China

Current position, grades: Experimentalist at Hebei Normal University of Science & Technology, Qinhuangdao.
University studies: Bachelor's degree in economics at Hebei Normal University of Science & Technology, Qinhuangdao, Hebei in 2008.
Scientific interest: Economics.
Publications: 1.

Operation Research and Decision Making

Model study on stochastic flow logistics network considering neighbourhood information of nodes

Dong-min Xu*, You-fang Huang, Bin Yang

Logistics Research Center, Shanghai Maritime University, 201306, China

Abstract

The significance and stochasticity of node have an effect on logistics network activity. This paper describes the importance of node by integrating the node neighbourhood information, which relates to the neighbourhood arc energy consumption and node degree which shows the node connectivity in logistics network, and during which this paper introduces the parameter α to represent the preference degree of energy consumption. This paper develops a multi-objective model seeking to energy consumption minimization and reliability maximization in the light of node stochasticity and the influence of node importance. The MPs (Minimal Path Set) is employed during the process of modelling. Genetic algorithm is applied to solve this model and a numerical experiment is presented to demonstrate the effect of significance and stochasticity of node to logistics path choice and flow assignment.

Keywords: stochastic flow network, neighbourhood information, energy consumption constraint, multi-objective model, logistics network

1 Introduction

Logistics network, as the supporter for logistics activity, not only ensures the implement of logistics activity but also affects the reality of logistics activity due to its structure specification and elements parameters. In general, the logistics network G is regarded as a network which composes of node/vertex set V and arc set A. And the network is noted $G = (V, A)$. V and A signify the supporter of logistics activity, and the network configure of $G = (V, A)$ has an important influence on logistics activity arrangement. The nodes undertake these activities such as loading and unloading, packaging, distribution processing and so on, and these nodes connect different edge/arc to fulfil the logistics activity. These properties of nodes, the difference on connectivity and stochasticity, affect the entire logistics network activity, on account of the presentation of node connectivity is significance of node, therefore this paper will do some research about impact analysis of the significance and stochasticity of logistics network to logistics activity.

The area of the node significance has been considered by many researchers who mainly focused on the measurement approach [1-10]. [1] proposed the node degree to appraise the importance of node. [3] presented betweenness centrality to evaluate the node importance. [2, 4] provided with the method of node deletion aiming at communications network system, and [4] described the significance of node adopting node contraction, and [6] improved the node contraction in weighted complex network. [5] improved the efficiency and validity of evaluating node importance based on node closeness and node key degree. [7] integrated the node degree, node betweenness and high peak hour traffic flow based on fuzzy clustering method to assess the node importance in traffic complex network. [9] presented the evaluation method of node importance base on load flow considering the contribution of the nodes for the whole network. [8] introduced the evaluation of edge importance considering the influence over node importance from the characteristics of node connection, and established the evaluation method of node importance based on agglomeration. [10] determined the node importance by devising the matrix of transfer efficiency and that of the operation influence coefficient, which considered the neighbourhood information of node for logistics network. [10] indicated that node importance may be different on account of information influence coming from node neighbourhood edge/arc and others nodes, although these nodes have the same node degree of node betweenness. All methods abovementioned give good results relating to some research area, however, it is not sufficient to logistics network, especially for new issue in logistics network area.

Logistics network, one form of stochastic flow network, has the properties of uncertainty and stochasticity deriving from node and/or arc service capability, and these properties influence entire logistics network activity arrangement. [11, 12] fully deals with the stochastic flow network, and indicated that each node and/or arc has a designated capacity. Nodes/arcs have different lower levels due to various partial and complete failures; therefore the system capacity is not fixed. To the question of stochastic-flow network, a majority of researchers concentrated on system reliability, especially the algorithm design of solving system reliability based on MPs [13-22] and MCs [11, 12, 23-27] in light of specific questions. In

addition a small number of researchers focused on the optimization of stochastic-flow network. [28] built a multi-objective model to minimize cost and maximize network reliability depending on the concept of MPs, and designed genetic algorithm to solve problem. [29] optimized the flow assignment of stochastic-flow logistics network considering the weight variety to attain these objectives of minimum cost, maximum network reliability, minimum maximal delivery time of single path. Similarity the MPs is introduced during the progress of modelling. The studies talked above displayed better results for reliability appraisal and solution, but only pay attention to reliability. Though certain researchers gave an emphasis on optimization problem of stochastic-flow network, they cannot give much attention to logistics network, in particular cannot integrate the stochasticity with node importance with regard to stringent pressure of energy-extensive consumption logistics network is facing.

With respect to energy saving and consumption reducing, it becomes a big question due to energy shortage and environmental pressure. Logistics is relatively energy-intensity and is rapidly expanding mainly as a result of trade increment. While energy consumption is related to the property of logistics network and parameters setting, this paper will demonstrate this aspect. So far a lot of researchers discussed energy consumption involved in logistics relevant activity. [30] built a framework with seven key parameters relating to the freight-intensity of economy, the division of freight traffic between modes, the utilization of vehicle capacity. The energy efficiency of logistics operations and the ratio of emissions to energy use, and then et author indicated that by altering these parameters companies and governments should be able to decouple the growth in demand for logistics from the associated energy requirements. [31] pointed out that the design of systems and processes in intra-logistics is an essential part of factory planning, and suggested that energy efficiency-oriented planning of logistics systems should be considered. [31] and [30] analysed the energy consumption in view of system perspective. [32] discussed the port logistics network activity and indicated that the energy consumption could be decreased by coordinating vessel berthing resource allocation and speed of ship. Undoubtedly the energy consumption have been crucial practical problems and attracted attention, however, the related work is shortage relatively. There is no research that discussed the influence of the properties of logistics network to logistics network energy consumption, thereby this paper will introduce the energy consumption factor into the appraisal system of node significance by virtue of node neighbourhood information and establish the relation between node significance and energy consumption.

According to the above analysis, this paper will analyse the influence of node importance considering energy factor and node stochasticity to logistics network activity, during which the paper redefine the node importance of stochastic logistics network considering neighbourhood information of node. And then a stochastic flow logistics network shall

be presented under energy consumption preference degree, during which the MPs is taken. Genetic algorithm is utilized to solve this problem and the numerical experiment demonstrated an effect of significance and stochasticity of node to logistics path choice and flow assignment as well as energy consumption.

2 Analysis of stochastic-flow logistics network (SFLN)

Logistics network, one form of stochastic-flow network, has common characters with the others stochastic-flow network such as power net and communication network. The common character is the capacity stochasticity of the network components (nodes and/or arcs) which followed some probability distribution coming from statistical information or empirical analysis. The uncertainty will act on the stochastic-flow network activities. Besides that, logistics network is a kind of complex network. Node significance is key issue in the area of complex network, therefore this paper shall give more attention to node significance of logistics network, which differentiate from general complex network. This section mainly refers to two sides: node stochasticity and node significance. Concerning node stochasticity, we deemed that capacity of node is a random variable, which follow certain probability distribution and impact the path choice and flow assignment for logistics activity. In respect to node significance, we proposed evaluation system of node significance which integrated into the energy consumption and indicated that the result of node significance have an effect on the logistics activities.

2.1 NODE STOCHASTICITY

Node stochasticity means that the capacity of node is not fixed but has different value at a certain probability, and the stochasticity leads to the uncertainty of logistics network activities. Assuming that the capacity of some node is an integer random variable, its distribution law is as follow:

$$\begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_k & \cdots \\ \hline p_k & p_1 & p_2 & \cdots & p_k & \cdots \end{array}, \; p_k \geq 0, \; k = 1, 2, \dots$$

and $\sum_{k=1}^{\infty} p_k = 1$, x_k is an integer.

We can see that distribution law of node capacity contains two sides context:

1) all possible value x_k,

2) p_k, probability corresponding to possible value. x_k and p_k jointly determine the capacity of node, and in further affect the path choice, flow assignment and logistics network transfer reliability.

All possible value of node capacity is defined as width domain, namely the interval length $[Min, Max]$; probability corresponding to possible value is defined as

depth domain, namely the value of p_k. In practice the width domain and depth domain of node capacity may change when alter logistics network structure, in other words the stochasticity of node capacity possesses changeability, and the changeable node capacity impact logistics network activities again.

We supposed that distribution law of node capacity is as follows in specific logistics network structure:

X	0	10	30	50	100
p_k	0.03	0.07	0.08	0.12	0.70

.

Once adjusting logistics network structure, the distribution law of node capacity may vary as follows:

X	0	20	50	80	120
p_k	0.02	0.08	0.15	0.20	0.55

.

In this paper we will not allow for the alteration of logistics network structure, so we only discuss the stochasticity of node capacity under given network structure without considering the alteration on distribution law. That means the logistics network structure is constant but not flexible.

2.2 NODE SIGNIFICANCE

Generally, the significance of node is closely related to the amount of neighbourhood arc, the more neighbourhood arc, the more flow bearing on node, and the class of node dominate the flowing of the entire logistics network activities; on contrary, the less neighbourhood arc, the less flow bearing, and this class of node effect small and lower importance. This method of appraisal to node significance is called node degree. In logistics network, the node significance has relation to the business relationship degree of neighbourhood connected node as well as node degree. And the relationship degree of neighbourhood connected node is interpreted as function difference of arc [10]. This paper will reconstruct the appraisal system of node significance based on flow bearing ability of node itself and business relationship degree of neighbourhood connected node we could also call neighbourhood information of node. The significance of node is composed of node degree which represents the connectivity of node and the weight of neighbourhood arc which is represented by energy coefficient reciprocal. Here the term "energy coefficient" is referred to energy consumption per unit flow when transporting from one node to another by some arc. Assumed that the energy coefficient matrix of the entire network is $E = \left[e_{i,j} \right]_{n \times n}$, $e_{i,j}$ represents the energy coefficient connecting node v_i and v_j, N represents the number of node. Then the weight of arc is showed:

$$w_{ij} = \begin{cases} 1/e_{ij} & i \neq j, v_j \in \Gamma_i \\ 0 & else \end{cases}, \tag{1}$$

where Γ_i represents the neighbourhood connected node set of v_i and $\Gamma = (\Gamma_1, \Gamma_2, ... \Gamma_n)$. Taking into consideration the inverse multi-arc between two connection nodes, such as a_{35}, a_{53} between nodes v_3, v_5, thereby Equation (1) is divided in:

$$\vec{w}_{ij} = \begin{cases} 1/\vec{e}_{ij} & i \neq j, v_j \in \Gamma_i \\ 0 & else \end{cases}, \tag{2}$$

$$\bar{w}_{ij} = \begin{cases} 1/\bar{e}_{ij} & i \neq j, v_j \in \Gamma_i \\ 0 & else \end{cases}, \tag{3}$$

where $\vec{e}_{ij}$ represents the energy coefficient of neighbourhood arc outflow of v_i, $\bar{e}_{ij}$ represents the energy coefficient of neighbourhood arc inflow of v_i. We define the restraint degree of node (RDN) deriving from neighbourhood arc energy coefficient is I_i, and I_i is expressed as follows:

$$I_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\vec{w}_{ij} + \bar{w}_{ij}) \quad v_j \in \Gamma_i. \tag{4}$$

The number of node v_i all neighbourhood arc is n_i. Normalization of I_i is expressed:

$$I_i^u = I_i / \sum_{i=1}^{n} I_i, \tag{5}$$

where n is the number of all nodes. Supposed that the node degree of v_i is d_i and the average degree of network is $\bar{\kappa}$. The node significance of v_i is indicated as following equation:

$$h_i = \alpha I_i^u + (1-\alpha)\frac{1}{n}\frac{d_i}{\bar{\kappa}}, \tag{6}$$

where I_i accounts for the restraint degree of node v_i flow from energy coefficient of node neighbourhood arc. If every arc which is connected with node v_i has high energy coefficient, the flow through (inflow and outflow) may be relatively low by way of v_i because of the energy consumption limitations for entire logistics network. The connection capability of v_i, that is d_i, reflect the absolute bearing capacity of node v_i in the whole logistics network. α is the weight. By adjusting the parameter of α, which expressed the emphasis for energy, we can fulfil the flow assignment based on the consideration of node connection capacity and logistics network energy consumption requirement. When the value of α is zero, it implies that the d_i of node degree totally dominated the bearing capacity of node v_i without considering the energy factor. When the value of α is one, it indicates that the I_i of

restraint degree of node v_i considering energy coefficient restriction of node neighbourhood arc dominated the bearing capacity of node v_i, and it means placing ultimate emphasis on energy for node flow through. By integrating I_i and d_i, we get the node importance of v_i, and then assign the node flow of v_i in according with its node importance considering the demand requirements.

Equation (6) shows that energy consumption is closely related to node importance. There are two sides, one side is the I_i of node restraint degree basically coming from $e_{i,j}$, energy coefficient of neighbourhood arc; the other side is the value of α, which indicates the preference strength for energy consumption. I_i, d_i, α all of them determined node importance. As a given logistics network, we assume I_i and d_i could be get immediately, and the value of α may vary depending on different preference strength in this paper.

3 Stochastic-Flow logistics network model

In this paper, we discuss flow assignment on arc and node for SFLN considering the node neighbourhood information which includes node significance and stochasticity. As far as node significance, we considered the influence of the energy consumption of node neighbourhood arc to node bearing capacity and further affect flow assignment; as far as stochasticity, we discuss capacity limitation and variation of node followed some probability distribution which causes the uncertainty for flow assignment of SFLN. We propose a multi-objective model for SFLN, and introduce *MPs* [14] (minimal path sets) to model. The objective of this model is energy minimization and reliability maximization. We represent SFLN as G in following discussion.

3.1 ASSUMPTIONS

- To develop a mathematical model, we first present the assumptions and notations respectively. The main assumptions considered in this problem formulation for the SFLN, are as follows.
- The capacity of each node is an integer-valued random variable which takes values according to a given distribution.
- The capacities of different nodes are statistically independent.
- Each arc is perfectly reliable without capacity limitation.
- Each node has only the ability of transhipment without considering energy consumption.
- There is no multi-arc with same direction between nodes; if there is a pair of arcs with inverse direction between two nodes, visiting the same node is not allowed in one path.
- Flow in SFLN must satisfy the so-called flow-

conservation.

3.2 PARAMETERS AND DECISION VARIABLES

The following notations are used in the model formulation.
Index sets:
$V = (v_i \mid 1 \le i \le n)$: set of nodes in G, n represents the number of transfer nodes.
$A = (a_{ij} \mid i = 1,...,n, j = 1,...,n, v_i, v_j \in V, i \ne j)$: set of arcs in G.
$S = (v_1, v_2,..., v_\sigma) \subset V$: set of source nodes, σ is the number of source nodes.
$T = (v_{n-\theta+1}, v_{n-\theta+2},..., v_n) \subset V$: set of terminal nodes, θ is the number of terminal nodes.
$N = V - \{S, T\}$: set of transfer nodes $v_m (m = \sigma+1,..., n-\theta)$.
$\Gamma_m \subset V$: set of neighbourhood connected nodes to transfer node $v_m \in N$.
$M = (M_m \mid m = \sigma+1,..., n-\theta, v_i \in N)$: set of capacity to transfer node, M_m is the maximum capacity to node $v_m \in N$ and the value of M_m is integer on an interval $[0, M_m]$ with some probability.
$U = (u_1, u_2,..., u_\sigma)$: set of quantity supplied to source nodes.
$D = (d_{n-\theta+1}, d_{n-\theta+2},..., d_n)$: set of quantity demanded to terminal nodes.
Ω_f : set of feasible solution.

Parameters of the model:
$x_m \in M_m (m = \sigma+1,... n-\theta)$: capacity of node $v_i \in N$ under certain condition.
$x_{fm} (m = \sigma+1,... n-\theta)$: flow by the way of transfer node $v_m \in N$ and x_{fm} should less than or equal to $x_m \in M_m$.
$X_f = (x_{f\sigma+1}, x_{f\sigma+2},... x_{fn-\theta})$: flow assignment for all transfer nodes under Ω_f.
f_{ij} : flow of arc a_{ij}, f_{ij} has not capacity limitation.
e_{ij} : energy consumption coefficient of a_{ij}.
$g(f_{ij})$: a concave function related to the flow of arc a_{ij} considering energy consumption.
$u_p (p = 1,..., \sigma)$: maximum supply to source node $v_p \in S$.
$d_q (q = n-\theta+1,..., n)$: demand of terminal node $v_q \in T$.
h_m : node significance of node $v_m \in N$, which influences the flow assignment of nodes.
$R(f)$: reliability of G satisfying Ω_f.
$E(f)$: total energy consumption of G on the condition of Ω_f.
$MPs_{p,q,k}$: the *k-th* path of *MPs* from source node v_p to terminal node v_q.

$k_{p,q}$: the number of *MPs* from source node v_p to terminal node v_q.

C_m represents the condition of $(x_{fm} \leq x_m)$.

$P(C_m)$ is the probability for the event satisfying $(x_{fm} \leq x_m)$.

Decision variables:

$f_{p,q,k}$ – flow assignment for some path from source node v_p to terminal node v_q. $f_{p,q,k}$ is not negative number and is integer. And $f = (f_{1,n-\theta+1,1}, \dots f_{p,q,k}, \dots, f_{\sigma,n,k_{\sigma,n}})$ is flow assignment for possible paths from source nodes to terminal nodes.

In addition:

- Minimal path means a path which consists of a series of successive nodes and arcs from a certain source node to a certain terminal node. All possible minimal paths constitute *MPs*.

- $MPs = \{MPs_{p,q,k} \mid p = 1, \dots, \sigma, q = n - \theta + 1, \dots, n, k = 1, \dots, k_{p,q}\}$ represent the set of all possible minimal paths.

The flow of transfer node $v_m \in N$ can be received using the following equation:

$$x_{mi} = \sum_{p=1}^{\sigma} \sum_{q=n-\theta+1}^{n} \sum_{k=1}^{k_{p,q}} \{f_{p,q,k} \mid v_m \in MPs_{p,q,k}, m = \sigma + 1, \dots, n - \theta\}, \quad (7)$$

As the significance of transfer node $v_m \in N$ has an effect to the flow bearing of transfer node $v_m \in N$, so we assumed that the flow bearing of each transfer node is in proportion to the value of its significance, and it can be expressed as below:

$$x_{fm} = \frac{\sum_{p=1}^{\sigma} \sum_{q=n-\theta+1}^{n} \sum_{k=1}^{k_{p,q}} \{f_{p,q,k} \mid v_m \in MPs_{p,q,k}\}}{\sum_{m=\sigma+1}^{n-\theta} x_{fm}} = h_m. \quad (8)$$

We could get f_{ij} of arc a_{ij} according to *MPs* that passed through arc a_{ij}. The formula is shown as the following:

$$f_{ij} = \sum_{p=1}^{\sigma} \sum_{q=n-\theta+1}^{n} \sum_{k=1}^{k_{p,q}} \{f_{p,q,k} \mid a_{ij} \in MPs_{p,q,k}\}. \quad (9)$$

3.3 MODEL

The reliability of entire SFLN depends on the condition $(x_{fm} \leq x_m)$ of every transfer node, because the capacity of transfer node is stochastic random variable and the capacity of every arc is not limited. In addition, we assumed theses capacities of different nodes are statistically independent, thereby we can show the reliability of SFLN as follows:

$$R(f) = \Pr\{C_{\sigma+1} \cap C_{\sigma+2} \cap \dots \cap C_{n-\theta}\} = \prod_{m=\sigma+1}^{n-\theta} P(C_m). \quad (10)$$

In respect to energy consumption, it contains energy consumption of every arc without node energy consumption according to assumption. The entire energy consumption is formulated as below:

$$E(f) = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} e_{ij} g(f_{ij}) \quad (a_{ij} \in Mps_{p,q,k}). \quad (11)$$

According to above analysis, we give the model. The objective function is expressed as following:

$$MaxR(f) \quad MinE(f) \quad \forall f \in \Omega_f. \quad (12)$$

Subject to:

$$\sum_{p=1}^{\sigma} \sum_{k=1}^{k_{p,q}} f_{p,q,k} = d_q, \quad q = n - \theta + 1, \dots, n \quad , \quad (13)$$

$$\sum_{q=n-\theta+1}^{n} \sum_{k=1}^{k_{p,q}} f_{p,q,k} \leq u_p, \quad p = 1, \dots, \sigma \quad , \quad (14)$$

$$\sum_{p=1}^{\sigma} \sum_{q=n-\theta+1}^{n} \sum_{k=1}^{k_{p,q}} \{f_{p,q,k} \mid v_m \in MPs_{p,q,k}, v_m \in N\} \leq M_m, \quad (15)$$

$$m = \sigma + 1, \dots, n - \theta,$$

$$f_{p,q,k} \geq 0, f_{p,q,k} \text{ is integer}. \quad (16)$$

In the proposed model of SFLN, the Equation (12) seeks to minimize the energy consumption and maximize the reliability in SFLN.

(*i*) the energy consumption associated with energy consumption coefficient of arc e_{ij} and the flow of arc f_{ij}.

(*ii*) the function $g(f_{ij})$ effect the total energy consumption of each arc considering the scale of flow.

(*iii*) the reliability associated with the reliability of every transfer node $P(C_m)$. Equation (13) ensures the demands of terminal nodes are met. Equation (14) guarantees the shipping quantity for every source node does not exceed its supply. Equation (15) bounds that flow of every transfer node less than the capacity limitation of every transfer node. Equation (8) enforces that the assigned flow bearing of every transfer node should be in apportion to its significance. Equation (16) enforces non-negativity and integer for the flow assignment of minimal path.

4 Numerical experiment

For a given SFLN, we need get the value of significance for every transfer node when given parameter α according to Equation (6) prior to model solution. In regard to others numerical value, such as node degree and $E = \left[e_{i,j} \right]_{n \times n}$, we could compute depending on definition and given data. Once getting the value of significance, we

can deal with the model.

The integer coded genetic algorithm will be adopted to solve the model in light of complexity. Supposed that the flow assignment $f_{p,q,k}$ of each minimal path is one chromosome and all minimal path set constitute a gene.

Penalty function is applied to deal with constraint and fitness function is determined by the ratio of reliability to energy consumption.

4.1 EXPERIMENT OF SFLN

To illustrate the applicability of the proposed model, a numerical experiment is conducted in this section for SFLN. The SFLN is given as following Figure 1, and the matrix of energy consumption coefficient for every arc is given in Figure 2.
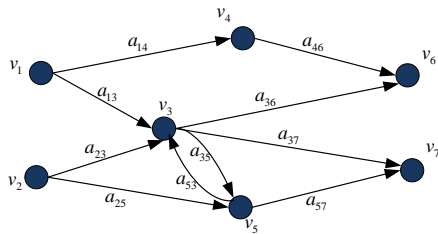
FIGURE 1 Stochastic-Flow Logistics Network

$$E = \begin{bmatrix} 0 & 0 & 8 & 2 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 8 & 4 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{7\times 7}$$

FIGURE 2 Matrix of Energy Consumption Coefficient

This experiment considers different value of parameter α to given SFLN, and we obtain the significance of transfer node and determine the principle of flow in accordance with significance of transfer node as follows (Table 1).

TABLE 1 Significance of transfer nodes on different preference and proportion of flow bearing

Significance of h_m parameter α	v_3	v_4	v_5	Proportion of flow bearing $v_3:v_4:v_5$
$\alpha =0.00$	0.50	0.16*	0.33	3:1:2
$\alpha =0.35$	0.44	0.22	0.33	4:2:3
$\alpha =0.65$	0.40	0.27	0.33	6:4:5
$\alpha =1.00$	0.34	0.33	0.34	1:1:1

Remarks:*The figures are rounded.

By search method, we get the nine minimal paths in Figure 1.

$MPs_{1,6,1} = \{v_1, a_{14}, v_4, a_{46}, v_6\}; MPs_{1,6,2} = \{v_1, a_{13}, v_3, a_{36}, v_6\};$

$MPs_{2,6,1} = \{v_2, a_{23}, v_3, a_{36}, v_6\}; MPs_{2,6,2} = \{v_2, a_{25}, v_5, a_{53}, v_3, a_{36}, v_6\};$

$MPs_{1,7,1} = \{v_1, a_{13}, v_3, a_{37}, v_7\}; MPs_{1,7,2} = \{v_1, a_{13}, v_3, a_{35}, v_5, a_{57}, v_7\};$

$MPs_{2,7,1} = \{v_2, a_{23}, v_3, a_{37}, v_7\}; MPs_{2,7,2} = \{v_2, a_{23}, v_3, a_{35}, v_5, a_{57}, v_7\};$

$MPs_{2,7,3} = \{v_2, a_{25}, v_5, a_{57}, v_7\}.$

The minimal path via transfer node is listed in Table 2.

TABLE 2 Minimal paths via transfer node

Minimal path, Transfer node v_m	$MPs_{p,q,k_{p,q}}$
v_3	$MPs_{1,6,2}, MPs_{2,6,1}, MPs_{2,6,2}, MPs_{1,7,1},$ $MPs_{1,7,2}, MPs_{2,7,1}, MPs_{2,7,2}$
v_4	$MPs_{1,6,1}$
v_5	$MPs_{2,6,2}, MPs_{1,7,2}; MPs_{2,7,2}, MPs_{2,7,3}$

Setting value: $U = (9,9)$, $D = (6,6)$, $g(f_{ij}) = \sqrt{f_{ij}}$, and we get the distribution law of capacity for transfer node as following Table 3.

TABLE 3 Distribution law of transfer node

Node	Capacity	Probability
v_3	0	0.000
	1	0.002
	2	0.003
	3	0.004
	4	0.006
	5	0.085
	6	0.125
	7	0.175
	8	0.200
	9	0.400
v_4	0	0.002
	1	0.003
	2	0.005
	3	0.010
	4	0.100
	5	0.145
	6	0.735
	7	0.000
	8	0.000
	9	0.000
v_5	0	0.001
	1	0.005
	2	0.009
	3	0.035
	4	0.075
	5	0.100
	6	0.785
	7	0.000
	8	0.000
	9	0.000

The population size is 200, and initialize. We get optimization flow assignment of all *MPs* after 300 iterations for parameter $\alpha = 1.00$. Part of the feasible values of objective function and corresponding code are given during iteration corresponding to Table 4. And the flow of each arc is get via Equation (9) corresponding to Table 5.

TABLE 4 Part of feasible values of objective function and corresponding code

$R(f)$	$E(f)$	$f_{1,6,1}$	$f_{1,6,2}$	$f_{2,6,1}$	$f_{2,6,2}$	$f_{1,7,1}$	$f_{1,7,2}$	$f_{2,7,1}$	$f_{2,7,2}$	$f_{2,7,3}$
7.67118E-01*	6.981360E+01	5	0	0	1	1	1	1	1	2
7.67118E-01	6.423753E+01	5	1	0	0	1	2	0	1	2
7.67118E-01	6.540765E+01	5	0	1	0	1	2	0	1	2
7.67118E-01	6.696739E+01	5	0	0	1	0	1	2	1	1
7.67118E-01	6.904259E+01	5	0	0	1	1	2	1	0	2
7.67118E-01	6.540765E+01	5	1	0	0	0	2	1	1	2
7.67118E-01	6.540765E+01	5	0	1	0	0	2	1	1	2

*Remarks: when $\alpha = 1.00$, the flow bearing of each transfer node is equal 1:1:1, that means each transfer node has the same flow bearing, thereby according to the Equation (10),we get the same value of reliability for different flow assignment. It is a special case.

TABLE 5 Flow of arc of feasible value corresponding to Table 4

f_{13}	f_{46}	f_{13}	f_{23}	f_{36}	f_{37}	f_{35}	f_{53}	f_{25}	f_{57}
5	5	2	2	1	2	2	1	3	4
5	5	4	1	1	1	3	0	2	5
5	5	3	2	1	1	3	0	2	5
5	5	1	3	1	2	2	1	2	3
5	5	3	1	1	2	2	1	3	4
5	5	3	2	1	1	3	0	2	5
5	5	2	3	1	1	3	0	25	5

Similarly, we get the optimal value for different parameter α .

TABLE 6 Optimal value to different preference

α	$M_{PS} = (f_{p,q,k})$	$f_{i,j}$	$R(f)$	$E(f)$
$\alpha = 0.00$	(3,0,1,2,1,2,1,2,0)	(3,3,3,4,3,2,4,2,2,4)	8.444400E+01	3.110860E-01
$\alpha = 0.35$	(4,0,0,2,1,2,1,2,0)	(4,4,3,3,2,2,4,2,2,4)	8.082551E+01	4.615800E-01
$\alpha = 0.65$	(4,1,0,1,1,1,1,1,2)	(4,4,3,2,2,2,2,1,3,4)	7.472573E+01	7.805700E-01
$\alpha = 1.00$	(5,1,0,0,1,2,0,1,2)	(5,5,4,1,1,1,3,0,2,5)	6.423753E+01	3.671180E-01

4.2 RESULT ANALYSIS

According to the result of Table 6, we draw the graph of Figure 3.
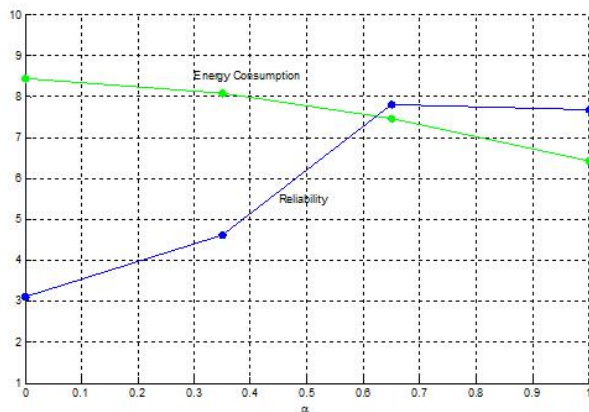
FIGURE 3 Comparison of optimal value for different preference

The horizontal axis represents parameter α of energy consumption preference; in vertical axis the green broken line represents energy consumption and the blue broken line is reliability. We can see that the value of energy consumption drops gradually along with the increase of parameter α , which can be explained that total energy consumption of SFLN may drop when given more emphasis to energy and then the emphasis finally effect the flow assignment by way of transfer node significance. Based on the experimental results, we do not find causal relation between parameter α and system reliability, and we think that it may be the result of stochasticity for transfer node. It should be studied in further. It is worth mentioning, because the significance of transfer node is introduce to influence flow assignment, the system reliability of SFLN must decrease compared with the same SFLN without restriction from significance of transfer node.

5 Conclusions

This paper has studied the SFLN considering neighbourhood information of transfer node. In light of energy consumption pressure to logistics activity, this paper proposed the restraint degree of node coming from neighbourhood arc energy coefficient and combined the node degree to represent the significance of node. The parameter α , which indicates the preference degree to energy is introduced to integrated RDN and node degree. For stochasticity of node, a bi-objective model has been developed and the significance of node embed in model had influence to flow assignment for model. The numerical experiments have shown that the energy consumption will drop along with the strength of preference to energy consumption. But the reliability of SFLN and parameter has no dependency according to present study.

Although this paper proposed the RDN and redefined the significance of node, and analysed the influence of α (actually via significance of node) to energy consumption,

we had not balanced the preference of energy and reliability, namely we had not attain the associativity between α and reliability. In addition, we discussed stochasticity of node in the given SFLN; however, the variation of distribution law to node has not been involved. And further steps for research will be extension of the SFLN.

Acknowledgments

References

[1] Callaway D S, Newman M E, Strogatz S H, Watts D J 2000 Network robustness and fragility: Percolation on random graphs *Physical review letters* **85**(25) 5468-71.

[2] Yong C, Ai-qun H, Jun H, Li-quan C 2004 A method for finding the most vital node in communication networks *Application Research of Computers* 2004 **14**(1) 21-4 *(in Chinese)*

[3] Budanitsky A, Hirst G 2006 Evaluating WordNet-based Measures Of Lexical Semantic Relatedness *Computational Linguistics* **32**(1) 13-47

[4] Yun-jin T, Jun W, Hong-zhong D 2007 Evaluation method for node importance based on node contraction in complex networks Systems Engineering – Theory & Practice **26**(11) 79-83 *(in Chinese)*

[5] Jing C, Lin-fu S 2009 Evaluation of node importance in complex network *Journal Journal Southwest Jiaotong University* **44**(3) 426-9 *(in Chinese)*

[6] Tao Z, Shui-ping Z, Rong-xiao G, Guo-cen C 2009 Improved evaluation method for node importance based on node contraction in weighted complex networks *Systems Engineering and Electronics* **31**(8) 1902-5 *(in Chinese)*

[7] Li W, Xin-yu Y, Ying-hong L, Zheng-xi L 2010 Traffic complex network node importance assessment based on fcm clustering *Journal of Transportation Systems Engineering and Information Technology* **10**(6) 169-73 *(in Chinese)*

[8] Jia-sheng W, Xiao-ping W, Wei L, Yong-qiang C 2012 Improved method of node importance evaluation in weighted complex networks *Computer Engineering* **38**(10) 74-6 *(in Chinese)*

[9] Hong-wei Y, Yong-Zhang, Huan-kun W, Yi-Liu 2013 New measure of node importance based on load flow in node-weighted complex networks *Application Research of Computers* **30**(1) 134-7 *(in Chinese)*

[10] Ying-yi H, Chun J 2013 Model and analysis for cascading failure on logistics network based on local information of nodes *Application Research of Computers* **30**(9):2625-8 *(in Chinese)*

[11] Lin Y-K 2002 Using minimal cuts to evaluate the system reliability of a stochastic-flow network with failures at nodes and arcs *Reliability Engineering & System Safety* **75**(1) 41-6

[12] Lin Y K 2003 Using minimal cuts to study the system capacity for a stochastic-flow network in two-commodity case *Comput Oper Res* **30**(11) 1595-607

[13] Lin J S, Jane C C, Yuan J 1995 On reliability evaluation of a capacitated – flow network in terms of minimal pathsets Networks **25**(3) 131-8

[14] Lin Y K 2001 A simple algorithm for reliability evaluation of a stochastic-flow network with node failure. *Comput Oper Res* **28**(13) 1277-85

[15] Lin Y K 2002 Two-commodity reliability evaluation for a stochastic-flow network with node failure *Comput Oper Res* **29**(13) 1927-39

[16] Lin Y K 2003 Extend the quickest path problem to the system reliability evaluation for a stochastic-flow network *Comput Oper Res* **30**(4) 567-75

[17] Lin Y K 2004 An algorithm to evaluate the system reliability for multicommodity case under cost constraint *Comput Math Appl* **48**(5-6) 805-12

[18] Lin Y K 2004 *IEEE Transactions on Reliability* **53**(3) 381-7

[19] Lin Y K 2009 Time version of the shortest path problem in a stochastic-flow network *Journal of Computational and Applied Mathematics* **228**(1) 150-7

[20] Chen J, Sun L-f 2009 Evaluation of node importance in complex network *Journal Journal Southwest Jiaotong University* **44**(3) 426-9

[21] Lin Y K 2010 System reliability of a stochastic-flow network through two minimal paths under time threshold *Int J Prod Econ* **124**(2) 382-7.

[22] Lin Y K 2010 *IEEE Transactions on Reliability* **59**(1) 2-10

[23] Lin Y K 2006 Evaluate the performance of a stochastic-flow network with cost attribute in terms of minimal cuts *Reliability Engineering & System Safety* **91**(5) 539-45

[24] Lin Y K 2007 On a multicommodity stochastic-flow network with unreliable nodes subject to budget constraint *Eur J Oper Res* **176**(1):347-60

[25] Lin Y K 2002 Study on the performance index for a multicommodity stochastic-flow network in terms of minimal cuts *Journal of the Chinese Institute of Industrial Engineers* **19**(3) 42-8 *(in Chinese)*

[26] Yeh W C 2004 A simple mc-based algorithm for evaluating reliability of stochastic-flow network with unreliable nodes *Reliability Engineering & System Safety* **83**(1) 47-55

[27] Doulliez P, Jamoulle E 1972 Transportation networks with random arc capacities *RAIRO-Operations Research-Recherche Opérationnelle* 1972 **6**(V3) 45-59

[28] Qiang L, Qing-zhen Z, Ming-feng H 2008 Study on multi-objective of flow control and allocation in stochastic-flow network *Computer Engineering and Applications* **44**(15) 94-6

[29] Xiao Z, Jin Z, Si-jing L, Qi X 2012 A method for flow allocation in sthochastic-flow logistics network based on variable weights *Journal of Highway and transportation Research and Development* **29**(007) 144-50

[30] McKinnon A C 2012 Reducing energy consumption and emissions in the logistics sector *Energy, transport, & the environment*: Springer 521-37

[31] Müller E, Hopf H, Krones M 2013 Analyzing energy consumption for factory and logistics planning processes. Advances in production management systems competitive manufacturing for innovative products and services: *Springer* 49-56

[32] Alvarez J F, Longva T, Engebrethsen E S 2010 A methodology to assess vessel berthing and speed optimization policies *Maritime Economics & Logistics* **12**(4) 327-46

Operation Research and Decision Making

Authors

Dong-min Xu, born in June, 1977, Tangshan, Hebei, China

Current position, grades: PhD student, logistics researcher center, Shanghai Maritime University, China.
University studies: Master Degree at Wuyi University in China.
Scientific interests: logistics network design and optimization, green supply chain management.
Publications: more than 5 papers.
Experience: 5 planning for enterprise, 3 municipal projects, 1 National Natural Science project.

You-fang Huang, born in June, 1959, Shanghai, China

Current position, grades: professor, logistics research center, Shanghai Maritime University, China.
University studies: PhD at Tongji university.
Scientific interests: supply chain management and logistics information system.
Publications: more than 90 papers and more than 10 works and copyrights.
Experience: over 70 national and provincial projects.

Bin Yang, born in November, 1975, Qingdao, Shandong, China

Current position, grades: professor, vice-director in logistics research center, Shanghai Maritime University, China.
University studies: PhD at Shanghai Jiaotong University.
Scientific interests: logistic information system.
Publications: more than 40 papers published 8 patents.
Experience: more than 10 major programs.

Research on the relationship between knowledge transfer models to various strategic alliances

Yang Zhao*, Jie Tan

School of Economics & Management, Beijing Jiao tong University, Beijing 100044, P.R. China

Abstract

As a part of knowledge management research, this paper focuses on knowledge transfer within strategic alliances and attempts to classify the basic models of knowledge transfer that actually take place within SA so that get to know on how partners chose the most suitable knowledge transfer model in SA. This paper presents four kinds of knowledge transfer model in strategic alliances by summing up the outcome of research on this issue. Based on the conclusion, this paper then discusses the correspondence between the types of the strategic alliances and choosing these knowledge transfer models and five corresponding propositions about the issue is given at last.

Keywords: knowledge management, strategic alliances, knowledge transfer, tacit knowledge, explicit knowledge

1 Introduction

As Francis Bacon said, "Knowledge is power". The power of knowledge is a very important resource for preserving valuable heritage, learning new things, solving problems, creating core competences, and initiating new situations for both individual and organizations now and in the future. How to manage this knowledge has become an important issue in the past few decades, and the knowledge management (KM) community has developed a wide range of technologies and applications for both academic research and practical applications. In addition, KM has attracted much effort to explore its nature, concepts, frameworks, architectures, methodologies, tools, functions, real world implementations in terms of demonstrating KM technologies and their applications. In modern business world, knowledge has emerged as the most strategically- significant resource of the firm [1]. This assertion characterizes well the recent research impetus centered on the role of knowledge-based resources in the firm and on competitiveness.

At the heart of the analysis of competitive advantage and its sustainability lies the issue of knowledge imitability. Accordingly, of all approaches to knowledge imitability between a knowledge holder and a knowledge seeker, strategic alliances (SAs) constitute perhaps the most adequate, but nevertheless challenging vehicle for internalizing the other's competency. Not surprisingly, the growing interest in how organizations learn from their partners and develop new competencies through SAs has led to the emergence of a distinct stream of research. That is, SA and knowledge transfer (KT) have been a focal point of recent international business research. The emerging literature on KT within SA has studied the process of KT

and imitability from different perspectives, including: how knowledge is managed in SAs [2,3]; how knowledge is transferred across partners [4]; how knowledge is acquired from the parents by the joint venture itself [5]; how knowledge about collaborating per se develops over time and impacts collaborative outcomes [6]; how knowledge impacts performance [4].

Despite this growing research on how KT works within SA, few researchers summarize the basic models of KT and general principles of choosing right models of KT. According to Inkpen [2], part of the problem has been the types of SA are various, and the process, affecting factors and criterion of performance evaluation of KT in each type of SA are different.

As a part of KM research, this paper focuses on KT within SA, and attempts to classify the basic models of KT that actually take place within SA in order to provide a clear focus on how partners chose the most suitable KT model in SA. This is a theoretical discussion based on extensive literature on the subject.

2 Theoretical model

The purpose of this paper is to examine how different resource types influence the choice of alliance forms. To assist us in this examination, in this section we identify two major types of resources and propose a four-part typology of alliance forms.

2.1 TYPES OF FIRM'S RESOURCES

Since firm resources are of various types, it is no surprise that scholars have proposed a number of resource typologies. These descriptive typologies, however, lack

---

Corresponding author- e-mail: zhaoy@bjtu.edu.cn

adequate theoretical underpinnings. Miller and Shamsie suggest that based on the notion of barriers to imitability, all resources may be classified into two broad categories: property-based resources (PBRs) and knowledge-based resources (KBRs) [7].

PBRs are legal properties owned by firms, including financial capital, physical resources, human resources, etc. Owners enjoy clear property rights to these resources, or rights to use the resources, so that others cannot take them away without the owners' consent. Thus, PBRs cannot be easily obtained, because they are legally protected through property rights in such forms as patents, contracts, and deeds of ownership [7]. Because others cannot take PBRs away, alliance partners will not be overly concerned about unintended transfers of these resources.

KBRs refer to a firm's intangible know how and skills. In contrast to PBRs, KBRs are not easily imitable owing to knowledge and information barriers. Others cannot easily copy or imitate KBRs, because they are vague and ambiguous. Thus, tacit know-how, skills, and technical and managerial systems not protected by patents, all fall in this category [8]. Imitating technological and managerial resources may be inherently "uncertain," because knowledge creation inevitably involves "irreducible ex-ante uncertainty" [9].

Besides imperfect imitability, technological and managerial resources are also imperfectly substitutable. Satisfactory substitutes and alternatives to superior technologies and managerial talents are often not available. Nevertheless, these resources are relatively mobile, because technologies and managerial talents may be acquired rather efficiently through the market. In contrast, organizational resources, such as culture and learning capacity, are deeply embedded in a firm and are thus characterized by imperfect mobility.

The key difference between PBRs and KBRs springs from the fact that the protection of knowledge barriers is not perfect [7]. Whereas PBRs enjoy near-perfect legal protection, PBRs are more vulnerable to unintended transfers. Once others get adequate access to PBRs, it is difficult to keep these resources within the confines of the firm for long. Consequently, alliance partners will be concerned with losing their KBRs through an alliance [10, 11].

2.2 TYPES OF KNOWLEDGE-BASED RESOURCES

KBRs can be classified in many different ways but some of the key dimensions that have previously been examined are:

- Individual, versus collective knowledge, based on levels of knowledge;
- Explicit, versus tacit knowledge, based on nature of knowledge.

Distinguishing the type of KBRs in a project is important because different types of KBRs have been shown to vary in their characteristics such as ease of transfer.

Individual knowledge is knowledge that can be wholly understood and retained by an individual. Collective knowledge is knowledge that is shared by a collective such as a team, an organization, an industry or a society [14]. It is also assumed in the organizational learning literature that individual learning is necessary but insufficient to produce organizational learning. It is also more than the sum of learning by individual members of the organization. For the occurring of organizational learning, knowledge must be accessible to others beyond individual learners and it must be subject to application, change and adaptation by others in the organization. In this paper, collective knowledge is further divided into team knowledge and organization knowledge for simplifying the description.

A second key distinction often made in the knowledge management literature is that between explicit knowledge and tacit knowledge. Explicit knowledge is that which can be readily stated and codified [12]. Tacit knowledge by contrast is difficult to state and can only be gained by experience or 'learning by doing' [12]. It is non-verbalized, intuitive, and unarticulated, and hard to communicate and deeply rooted in action, involvement and commitment within a specific context. It is "a continuous activity of knowing" [13]; it is "the way things are done around here" [14]. Tacitness assumes that individuals know more than they can tell and tacit knowledge is often context specific and has a 'personal quality'. It would also involve providing opportunities for the teams to put the knowledge into action, either through role-playing or case-related activities, to allow for the type of tacit–explicit conversions [13].

This classification is essential as it provides a context through which our understanding of the effect of knowledge value on the form of SA can be furthered. Theoretically, Reed and DeFillippi singled out tacitness as a key source of ambiguity that raises barriers to imitation [12]. Empirically, in their study of the transfer of manufacturing capabilities, Zander and Kogut found that, indeed, the degree to which capabilities are modifiable and teachable (i.e., are non-tacit) significantly influences the speed of their transfer [15].

2.3 APPROPRIATION CONCERNS AND PARTNER PROTECTIVENESS

In an exploration of literature on partner appropriation concerns: Gulati posits that firms entering SA face two primary moral concerns; first, the unpredictability of the behaviour of partners, and second, the costs to a firm if a partner engages in opportunistic behaviour [16]. Indeed, prior research on the selection of forms of SA has been influenced primarily by transaction cost economics, which holds that organizations and managers will act opportunistically to maximize profit even at the expense of another party. Knowledge leakage is an important form of appropriation and therefore partner protectiveness is adopted, which influences the effects of KT [17].

In SAs, the protection of proprietary knowledge from partners is a vital issue to many firms [2, 6]. Transferring partners must have an incentive to palliate the cost typically associated with the transfer [18]. If not, partners can adopt explicit measures, deploy shielding mechanisms, and engage in defensive actions to protect the transparency of their competencies, particularly when the embodied knowledge is explicit and held by only a few experts [10, 19]. Hence, protection of technological know-how is likely to be prevalent and actively managed. Therefore, partner protectiveness is expected to lead to directly impede KT. Thus, in this study, associate with to knowledge tacitness, which is also expected to exert a direct effect on the transfer outcome, partner protectiveness is regarded as the main factor.

2.4 FORMS OF STRATEGIC ALLIANCES

A SA is defined as a long-term cooperative arrangement between two or more independent firms that engage in business activities for mutual economic gain, and an international SA is one where partners come from two or more countries. The role of alliance form has been argued that the extent of KT among partners is likely to depend on the alliance form [18].

SAs can take a variety of forms, including, but not limited to joint ventures, minority equity alliances, R&D contracts, joint R&D, joint production, joint marketing and promotion, enhanced supplier partnership, distribution agreements, and licensing agreements. Most studies on alliance structural choice have been based on the dichotomy of equity alliance vs. non-equity alliances [20]. Whereas equity alliances include equity joint ventures and minority equity alliances, non-equity alliances refer to all other cooperative arrangements that do not involve equity exchange. For non-equity alliances, Mowery et al. suggest two types: unilateral contract-based; and bilateral contract-based [11]. Integrating the above approaches into the classification of alliance structures, we adopt the following four part alliance typology: equity joint ventures (EJVs); minority equity alliances (MEAs); bilateral contract-based alliances (BCAs); and unilateral contract-based alliances (UCAs).

In an EJV, firms similarly pool their resources, but also create a new entity that is jointly owned and operated by two or more allying firms. The new entity is created to substantially integrate the joint efforts of partners in which the partners literally work together. One key problem in EJVs is that firms may be opportunistic in maximizing their own particular interests, to the detriment of their partners, such opportunistic behaviour tend to be more severe when it involves tacit knowledge and skills that are not protected by property laws.

In MEAs, one or more partners take all equity position in others Das and Teng argue that shared ownership helps control opportunistic behaviours. Since equity arrangements are rather complicated to implement as well as to get out of, they are usually entered into for longer

time horizons compared to alliances without equity investments [21]. A long duration for an alliance provides an incentive to partners to behave honestly and curb opportunistic behaviour.

UCAs embody a well-defined transfer of property rights, such as the "technology for cash" exchange in licensing agreements. Licensing, distribution agreements, and R&D contracts are the main forms of UCAs. The key feature here is that individual forms carry out their obligations independently of others. Such contracts tend to be complete and specific, and partners are expected to perform on then own accordingly, without much coordination or collaboration. Thus, the level of integration is relatively low in UCAs [11].

On the other hand, BCAs emerge when the partners have sustained production of property rights. BCAs require partners to put in resources and work together on a continuing basis for the purposes of collaboration but do not form a separate legal entity. Joint R&D, joint marketing and promotion, joint production, and enhanced supplier partnership are some good examples of BCAs [11].

As compared to unilateral contracts, bilateral contracts are usually incomplete and more open ended. To some extent, partners of UCAs have to let the cooperative relationship unfold itself.

3 Research model

Contractor and Ra posited that with respect to the appropriation concerns of the knowledge supplier; if knowledge is deeply embedded or tacit, then a partner does not easily copy such knowledge and therefore fears of opportunism are lower [22]. Conversely, if knowledge is codified or easily observable, then the knowledge supplier's appropriation concerns will be high and this may possibly lead to more hierarchical forms of governance. It is assumed that for an alliance to be formed there must be an exchange of knowledge or information (which can be viewed as codified or easily understood knowledge), be it technical, organizational, or policy data. Therefore, the nature of the knowledge exchanges needs to be considered at the first.

In revisiting the question raised by Contractor and Ra's model [22] regarding the manner in which the tacitness of the knowledge supplier influence the choice of forms of SA, it emerges from the literature reviewed that although are of significance in the choice of governance form, other factors such as the partner protectiveness and bargaining abilities of partner firms are also significant.

The research model for this paper draws on knowledge-based and transaction cost theories, Contractor and Ra's theoretical model and several conceptual models related to KT at different levels of analysis [23] and within different forms of SA [5]. It attempts to integrate views from interrelated research as it concerns the various types of resources which is transferred must consist with the suitable forms of SA in order to ensure success. Figure 1

illustrates the primary constructs and relationship of interest in this paper.

FIGURE 1 Research model

4 Notation basic model of KT within SA

We now discuss basic models of KT in terms of the four major categories of alliances outlined above: EJVs, MEAs, BCAs, and UCAs.

4.1 EQUITY JOINT VENTURES

Among various alliance forms, EJVs are the most instrumental in the transfer of tacit and explicit knowledge between the partners, because of the significant extent to which partners are exposed to each other [24]. Hamel also holds that when the partners work shoulder to shoulder in the same entity for an extended period, it becomes difficult to keep others from accessing one's tacit know-how [10]. Consequently, EJVs provide the best opportunities to acquire partner's KBRs (especially tacit ones). Researchers note that partners often use alliances as a cover for appropriating KBRs [19]. By the same token, Contractor and Ra believe that the transfer of tacit knowledge necessitates longer-term alliances with higher levels of partner interaction [22]. Von Hippel also holds that the transfer of tacit knowledge requires more face-to-face and longer personal contact between companies [25]. Moreover, Contractor and Ra hypothesized that the greater the complexity of knowledge, the more likely it is that EJV arrangements would be adopted by the allies [22]. This proposition is congruent with transaction cost arguments which propose that because complex technologies are more valuable, the associated consequences of opportunism are higher, and therefore EJVs (which are less reversible) is likely to be chosen as a form of SA [22]. Mowery et al. found that EJVs significantly facilitated inter-film transfer of technologies, resulting in greater technological similarities between the partners [11].
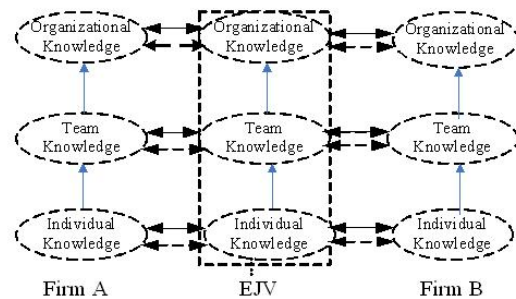
FIGURE 2 KT model in EJVs

However, the closer interaction of partners have, the more likely knowledge leakage (KL) would occur. Seen from Figure 2, KL (shown by the broken line, same in the followings) would emerge at all the levels, which is reluctant for the parents firm to accept because excessive core KL would damage the interests of the parents firm and leads to the termination of the alliance in the end. Hence, in this model of KT, the responsibilities of knowledge gatekeeper are very important for the parents firm.

Consequently, EJVs provide the best opportunities to acquire partners' KBRs including tacit & explicit knowledge while KL is a problem to parents firm at the same time.

4.2 MINORITY EQUITY ALLIANCES

Seen from Figure 3, in MEAs, the channel of knowledge learning is set up by sharing each other's ownership. Comparing to EJVs, there's no medial entity of joint venture, and controlling to the partner is enforced by taking its equity position.
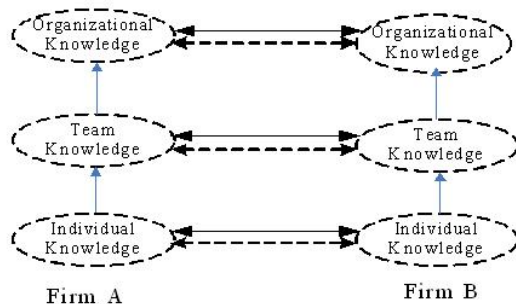
FIGURE 3 KT model in MEAs

In this model, KT is carried out under the supervision of all the partners in SA. The effect and efficiency of KT is impacted directly by the learning abilities of the supervisors of the partner companies. It is easy to deduce that the transfer of explicit knowledge is easier while tacit knowledge is more difficult for lack of close interaction within respective partners. On the other hand, KL is not avoidable but the partner could reduce the leakage of its core knowledge efficiently through strict control to its supervisor.

4.3 BILATERAL CONTRACT-BASED ALLIANCES

Because EJVs facilitate the process of transferring KBRs, they can be a disadvantage if both partners have substantial KBRs in an alliance. Thus, EJVs may be too risky a choice in such situations. Hence, contract-based alliance is better choice.

In BCAs, knowledge transferred within partners is mainly explicit, which is stipulated in the contract. At the same time, little tacit knowledge would be transferred. However, in some situations, the respective resources of partners (including but not limited explicit knowledge) which are stipulated in the contract too would be put into co-operations of alliance, which may result in the transfer of tacit knowledge to some extent. Seen from Figure 4, in this model, KT is limited rigidly by the contract and explicit knowledge is mainly the object of both partner's learning. So, little core knowledge would be leaked because of the existent contract.
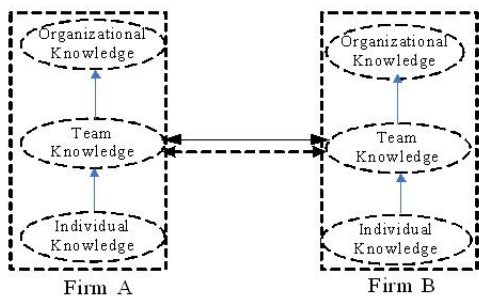
FIGURE 4 KT Model in BCAs
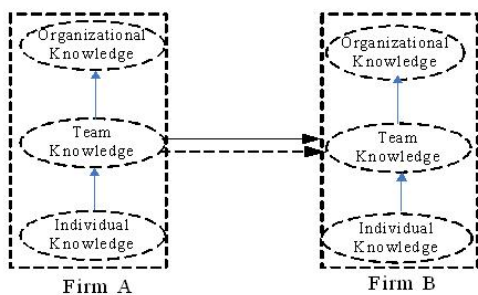
FIGURE 5 KT Model in UCAs

4.4 UNILATERAL CONTRACT-BASED ALLIANCES

Mowery et al. suggested inter-firm knowledge transfers should be more limited in UCAs such as licensing agreements, and postulate that UCAs should create fewer opportunities for inter-firm knowledge transfer. Thus, they determined that EJVs appeared to be more effective conduits for the transfer of complex capabilities than contract-based alliances such as licensing agreements [11]. Furthermore, Mowery et al. concluded that lower levels of transfer occur in unilateral contracts than in bilateral non-equity arrangements [11], while Oxley found that EJV as an alliance structure outperforms alternatives in supporting inter-firm learning [26].

Seen from Figure 5, in this model, KT is very similar to what in BCAs. The difference is that the direction of KT is unilateral and there is no feedback. KL is little in the process of KT hence the partner firms do not worry about KT.

5 Choice of KT model

Seen from the above-mentioned analysis, KT and KL are usually companied in spite of which KT model is adopted while the likelihood and the extent of KL in different KT are not in the same. In order to simplify the issue, two enterprises are supposed to be the partners in SA and KT within SA is therefore to be regarded as transferring between the two partners. Combined with the above classification of resources attributed to SA, 9 cases are deduced and shown by Table 1. According to the different case which Table 1 showed, we then analyse the choice of KT models.

TABLE 1 Matrix of Resources Attributed to SA

Firm B \ Firm A	Tacit	Explicit	Property
Tacit	1	2	3
Explicit	4	5	6
Property	7	8	9

Given firm A and firm B is of no difference, Case 2,3,6 is the same with 4,7,8 respectively, so we just discuss Case 1,2,3,5,6,9.

Case 1: Both sides mainly attribute tacit knowledge and want to acquire tacit knowledge too from the partner. Because an EJV is primarily a device to obtain tacit knowledge based on the above KT model analysis. Here is the following proposition:

P1: Both partner firms will prefer a EJV if, in the prospective SA, both of them attribute primarily tacit knowledge and want to acquire primarily tacit knowledge too from its partner.

Case 2: Firm A mainly attributes explicit knowledge while firm B mainly attribute tacit knowledge. To firm A, it want to acquire the tacit knowledge from its partner of course, so it prefer an EJV. To firm B, which should attribute tacit knowledge, it wants to just acquire its partner's explicit knowledge and at the same time, it has to prevent KL. In this case, EJV will not be preferred, for two reasons. First, there are no substantial KBRs contributed by the partners available for exploitation. Second, there are altogether too much of firm B's own KBRs that the partner could potentially appropriate, making it too risky to form an EJV. Contract-based alliance will also be less attractive to firm B in such case, because it doesn't offer sufficient safeguards against opportunistic behaviour regarding KBRs. Should a partner be found appropriating others' KBRs to an undue extent, its equity stake may be held as hostage. Thus, equity investments provide some protection against the unintended transfer of partners' tacit knowledge [21]. Hence, a MEA will be the preference for firm B. Then there is a conflict when KT model is chosen. The solution depends on the bargaining ability of the partners

and would be dealt with in the process of negotiation. Here's the following proposition:

P2: When firm A mainly attribute explicit knowledge while firm B mainly attribute tacit knowledge. The former prefer an EJV while the latter prefer a MEA, and the final KT model is determined by the bargaining ability of the partners.

Case 3: Firm A mainly attributes PBRs while firm B mainly attribute tacit knowledge. Then, firm A will prefer EJV to acquire tacit knowledge and will not wary about losing its own PBRs in a highly integrated operation characteristic of an EJV. This is because PBRs are protected by property rights, minimizing the likelihood of unintended transfer of resources. To firm B, which want to acquire PBRs and would attribute tacit knowledge, the most important thing is to prevent KL. That means it wants to stop its partner's learning process to its own tacit knowledge as soon as possible. Scholars suggest that once learning has been accomplished, alliances are likely to be intentionally terminated [19]. Hence, contract-based alliances, which are much easier to dissolve, will be preferred over EJVs and MEAs. Here is the following proposition:

P3: When firm A mainly attribute PBRs while firm B mainly attribute tacit knowledge. The former prefer an EJV while the latter prefer a contract-based alliance, and the final KT model is determined by the bargaining ability of the partners.

Case 5: Both sides mainly attribute explicit knowledge and want to acquire explicit knowledge too from the partner. Both of them will chose the KT model, which best facilitates the transfer of explicit knowledge and at the same time, prevents KT of tacit knowledge of their own. Based on the analysis about BCAs, here is the proposition:

P4: Both partner firms will prefer a BCA if, in the prospective SA, both of them attribute primarily explicit knowledge and want to acquire primarily explicit knowledge too from its partner.

Case 6: Firm A mainly attributes PBRs while firm B mainly attribute explicit knowledge. Based on the analysis about BCAs, here is the proposition:

P5: When firm A mainly attribute PBRs while firm B mainly attribute explicit knowledge, both partner firms will prefer a UCA.

Case 9: It is of no need to discuss in this case because both sides in SA mainly attribute PBRs, therefore KT won't occur and no model of KT should be considered to be chosen.

6 Conclusions

Although the concepts of SA and KT are gaining wide currency among both academics and practitioners, the theoretical link between the two is still weak. In this article 2 type of knowledge, 4 forms of basic SA models have been identified and 5 untestable propositions suggested. In particular, it has introduced knowledge tacitness as a key determinant of choosing KT models but it is also fundamental to consider the role of other factors such as KL and bargaining abilities. It has been demonstrated in the end that the types of knowledge which will be attributed and acquired by the partner firms within SA and their bargaining abilities are the main factors which impact the choice of KT models in SA.

Managers would also benefit from the discussion. As they understand more about the nature of the SA & KT, they can chose the right mode of SA and plan their cooperation strategies more effectively.

The aim of this study was to advance our understanding of the process of KT across alliance partners at both the conceptual and theoretical levels through some propositions. These propositions shed some important light on the mechanisms that facilitate or hinder KT between alliance partners. However, this paper does not offer an empirical test, which is the limitation of it.

References

[1] Grant R M 1996 Prospering in Dynamically Competitive Environments: Organizational Capability as Knowledge Integration *Organization Science* **7**(4) 375–87

[2] Inkpen A C 2002 Learning, Knowledge Management, and Strategic Alliances: so many Studies, so many Unanswered Questions *Cooperative Strategies and Alliances F J Contractor and P Lorange eds. Elsevier Science Amsterdam* 267–89

[3] Martin X, Salomon R 2002 Knowledge Transfer Capacity and Cooperative Strategies: When should Firms Leverage Tacit Knowledge through Licensing and Alliances? *Cooperative Strategies and Alliances F J Contractor and P Lorange eds Elsevier Science Amsterdam* 729–48

[4] Appleyard M M 1996 How does Knowledge Flow? Inter-firm Patterns in the Semiconductor Industry *Strategic Management Journal* **11** 137–56

[5] Lyles M A, Salk J E 1996 Knowledge Acquisition from Foreign Parents in International Joint Venture *Journal of International Business Studies* **27**(5) 877–903

[6] Simonin B L 1999 Ambiguity and the Transfer of Knowledge in Strategic Alliances *Strategic Management Journal* **20** 595–623

[7] Miller D, Shamsie J 1996 The Resource-based View of the Firm in Two Environments *Academy of Management Journal* **39** 519–43

[8] Hall R 1992 The Strategic Analysis of Intangible Resources *Strategic Management Journal* 9 135–44

[9] Lippman S, Rumelt R 1982 Uncertain Limitability: an Analysis of Inter-firm Differences in Efficiency under Competition *Bell Journal of Economics* 13 418–38

[10] Hamel G 1991 Competition for Competence and Inter-partner Learning within International Strategic Alliances *Strategic Management Journal* 12 (special issue) 83–103

[11] Mowery D C, Oxley J E, Silverman,, B S 1996 Strategic Alliances and Inter-firm Knowledge Transfer *Strategic Management Journal* 17 77–91

[12] Reed R, De Fillippi R J 1990 Causal Ambiguity, Barriers to Imitation, and Sustainable Competitive Advantage *Academy of Management Review* 15 88–102

[13] Nonaka I 1984 A Dynamic Theory of Organizational Knowledge Creation *Organization Sciences* 10 14–37

[14] Spender J C 1996 Making Knowledge the Basis of a Dynamic Theory of the Firm *Strategic Management Journal* 17 (special issue) 45–62

[15] Zander U, Kogut B 1995 Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: an Empirical Test *Organization Sciences* **6**(1) 76–92

[16] Gulati R 1998 Alliances and Networks *Strategic Management Journal* 19 293–317

[17] Beamish P W, Killing J P 1997 Cooperative Strategies: North American Perspectives 3ed San Francisco The New Lexington Press 2 68–73

[18] Dyer J, Singh H 1998 The Relational View: Cooperative Strategy and Sources of Inter-organizational Competitive Advantage *Academy of Management Review* **23**(4) 660–79

[19] Inkpen A C 1997 An Examination of Knowledge Management in International Joint Ventures *Cooperative Strategies and Alliances P Beamish and J Killings eds North American Perspectives. San Francisco New Lexington Press CA* 337–69

[20] Osborn R N, Baughn C C 1990 Forms of Inter-organizational Governance for Multinational Alliances *Academy of Management Journal* 33 503–19

[21] Das T K, Teng B S 2000 A Resource-based Theory of Strategic Alliances *Journal of Management* **26**(1) 31–61

[22] Contractor F J, Ra W 2002 How Knowledge Attributes Influence Alliance Governance Choices: a Theory Development Note *Journal of International Management* 8 11–27

[23] Argote L, Ingram P 2000 Knowledge Transfer: a Basis for Competitive Advantage in Firms *Organizational Behavior and Human Decision Processes* **82**(1) 150–69

[24] Kogut B, Zander U 1993 Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation *Journal of International Business Studies fourth quarter* 625–45

[25] Hippel E V 1994 Sticky Information and the Locus of Problem Solving: Implications for Innovation *Management Science* **40**(4) 429–39

[26] Oxley J E 1997 Appropriability Hazards and Governance in Strategic Alliances: a Transaction Cost Approach *Journal of Law, Economics and Organization* **13**(2) 387–409

Authors

Yang Zhao, born on December 7, 1973, Jinan, Shandong, P.R. China

Current position, grades: lecturer of Beijing Jiao tong University.
University study: Ph.D. degree in Business Management, School of Economics & Management, Beijing Jiao tong University, 2009.
Scientific interest: mathematics, stochastic processes, pattern recognition, digital statistical decision theory, adaptive control.
Publications: 3 papers.

Jie Tan, born on April 9, 1975, Shaoyang, Hunan, P.R. China

Current position, grades: lecturer of Beijing Jiao tong University.
University studies: Ph.D. degree in industrial economics, School of Economics & Management, Beijing Jiao tong University, 2011.
Scientific interest: mathematics, stochastic processes, employment, digital statistical decision theory, adaptive control.
Publications: 1 paper.

Power transformer diagnostic prediction research based on quantum neural networks and evidence theory

Qiang Song[1*], Ai-min Wang[2]

[1]*School of Mechanical Engineering, Anyang Institute of Technology, Henan, Anyang City 455000, China*

[2]*School of Computer and Information Engineering, Anyang Normal University, Henan, Anyang City 455000, China*

Abstract

Aiming at the fault of power transformer fault information diversity and uncertainty, a large amount of data and no regularity characteristics, a new fault diagnosis method of quantum neural network based on information fusion. In order to accurately and effectively identify transformer fault model, combining the quantum neural network and evidence theory combination of transformer fault diagnosis. A quantum neural networks to collect data on the macroscopic, microscopic quantum corrections in the interval of fuzzy intersection data according to a certain proportion of the rational allocation of the associated mode, so as to improve the accuracy of pattern recognition; use of the evidence theory can improve the convergence speed of quantum neural networks. The results were compared with the diagnosis and BP neural network input, that this method has a higher accuracy in transformer fault pattern recognition.

Keywords: transformer diagnostic prediction, combination forecasting, quantum neural networks, evidence theory

1 Introduction

Transformer plays an essential role in power system. A lack of coherent maintenance strategy may result in accidence of power system and reduce the reliability. Online monitoring for power transformer is important for safe operating. In order to keep transformer in good condition, diagnosis has become increasingly essential [3-5]. Although there are different methods used for detecting transformer fault. Dissolved gas in transformer oil is widely used as a reliable approach. In this paper, we present the RBF networks with self-adjustable number of hidden neurons for transformer faults detection. A new type of SOM RBF network together with its training algorithm is proposed [6, 7].

As a result, this has made the implementation of an effective neural network used in transformer fault detection system easy. The advantages of the proposed RBF neural networks are twofold. First, the best possible network architecture is determined by E-mail: author@domain.com the proposed training algorithm according to the input data. It does not require many trial tests. Second, the outputs of the neural network are able to not only perform fault detection, but also indicate the extent of the fault. In the first advantage, a cell-splitting grid (CSG) neural network, which is an extended SOM, is used to automatically determine the centres and the number of hidden neurons of the RBF networks. After completion of the training, the learned network is able to detect different types of faults. Our obtained results indicate that the proposed neural network approach is promising for diagnosing transformer faults via analysing the dissolved gases in transformer.

Transformer is an indispensable equipment of power system, whose fault state directly affects power supply reliability and system's operation. As the power network develops towards highly automation, growing demand has been put forward to the liability of power system. It's an urgent requirement to improve present equipment maintenance system. Therefore, increasing attention is being paid to state maintenance system based on online condition monitoring and fault diagnosis. It is a trend for this state maintenance system to replace the preventive maintenance system. Measurement and studies show that, the use of gas chromatography (Dissolved Gas Analysis) for transformer latent faults analysis is one of the most effective measures to ensure transformer operation, which has been the most effective way for equipment fault diagnosis at home and abroad [1]. Currently, Improved three-ratio method (formerly improved electrical committee agreements) is the main method to diagnose internal faults of transformer.

Among the various transformer fault diagnosis measures, dissolved gas analysis (DGA) is based on the principle that different types of transformer faults correspond to different dissolved gas concentrations. It detects transformer's latent fault by analysing the concentration of fault characteristic gases (H_2, CH_4, C_2H_2, C_2H_6, C_2H_4 etc.). Besides, DGA method can make the diagnosis under energized condition. Thus, it can periodically make internal diagnosis of transformer during its operation, without interference from external

---

Corresponding author e-mail: songqiang01r@126.com

electromagnetic field. So this method has been commonly and widely used in fault diagnosis in electrical equipment's [2].

2 Quantum neural networks

Quantum Neural Network appeared in 1990s. With the idea of quantum mechanics introduced into neural network study, it can overcome the flaw and deficiency of conventional neural network. Quantum neural network is the extension traditional neural network. By taking some advantages of quantum computations, parallel computation especially, quantum neural network has more parallel processing power, capable of handling much larger data sets. Therefore, quantum neural network has unprecedented advantages in data processing.

Currently, quantum neural network has been applied to fault diagnosis for power electronic circuit [6], soft fault diagnosis of tolerance analogue circuits [7], speech noise reduction [7] and so on. Compared with traditional neural network, quantum neural network has the following advantages: Set the input value of the system as $X = (X_1, X_2, ..., X_n)$, output value as $Y = (Y_1, Y_2, ..., Y_n)$. The action function of three-layer σ is Sigmoid function, and its weights and number neurons intervals, $\theta_v (v = 1, 2, 3, ..., ns)$ is quantum interval. Its size is the same as the number of fault modes to be diagnosed, namely, the same as the number of fault components. *sf* is the steepness factor.

a) Exponential memory capacity and memory speed [8,9];

b) high-speed learning and information processing capabilities;

c) capable of eliminating catastrophic amnesia, with no mutual interference between modes.

High stability and reliability. Quantum neural network is the combination of quantum computation and traditional neural network. Generally, there are two integrated forms:

a) to introduce quantum computing theory in the structure and training process of neural networks;

b) to design topology and training algorithm of neural network by directly borrowing some principles and definitions of quantum theories.

The structure and learning algorithm of quantum neural network are described below.

2.1 STRUCTURE OF QUANTUM NEURAL NETWORK

The structure of quantum neural network is shown is Figure 1, which includes the input layer, the first hidden layer, the second hidden layer and the output layer respectively $\omega1_{k,n}$, $\omega2_{m,k}$, $\omega3_{s,m}$, K, M, S.
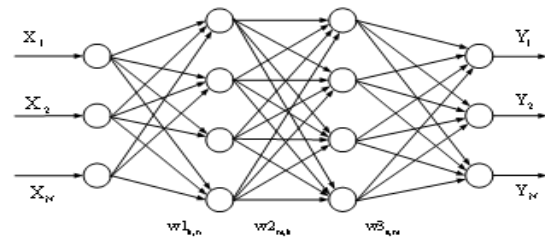
FIGURE 1 The structure of four-layer quantum neural network

Based on its network structure, the network output is:

$$Y_s = \sigma \left(\sum_{n=1}^{N} \omega3_{s,m} \sigma \left(\sum_{k=1}^{K} \left(\begin{array}{c} \omega2_{m,k} \times \dfrac{1}{ns} \times \sigma \\ \left(sf * \sum_{n=1}^{N} \omega1_{k,n} x_n - \theta_v \right) \end{array} \right) \right) \right), \quad (1)$$

where $s = 1, 2, 3, ..., S$; $k = 1, 2, 3, ..., K$; $m = 1, 2, 3, ..., M$. Set the total number of input samples as P, output expected value as $\hat{Y}_s^p$, based on Equation (1), the *s-th* node, *P-th* sample of network output layer is note as Y_P^S, then the error energy function expression can be defined as:

$$e = \frac{1}{2} \sum_{s=1}^{S} \sum_{p=1}^{P} \left(\hat{Y}_s^p - Y_s^P \right)^2, \quad (2)$$

where the partial derivative of *e* with respect to $\omega1_{k,n}$, $\omega2_{m,k}$, $\omega3_{s,m}$ can be derived as:

$$\frac{\partial e}{\partial w3_{s,m}} = -\sum_p \left(Y_S^P - \hat{Y}_S^P \right) \times \sigma' \times$$
$$\sigma \left[\sum_{k=1}^{K} w2_{m,k} \times \frac{1}{ns} \times \sigma \left[sf \sum_{n=1}^{N} w1_{k,n} x_n - \theta_v \right] \right], \quad (3)$$

$$\frac{\partial e}{\partial w2_{m,k}} = -\sum_p \sum_s \left(Y_S^P - \hat{Y}_S^P \right) \times \sigma' \times w3_{s,m}$$
$$\times \sigma' \times \frac{\sigma}{ns} \left[sf \sum_{n=1}^{N} w1_{k,n} x_n - \theta_v \right], \quad (4)$$

$$\frac{\partial e}{\partial w1_{k,n}} = -\sum_p \sum_s \sum_m \left(Y_S^P - \hat{Y}_S^P \right) \times \sigma'$$
$$\times w3_{s,m} \times \sigma' \times w2_{m,k} \times \frac{sf}{ns} \times \sigma' x_n^P . \quad (5)$$

2.2 LEANING ALGORITHM OF QUANTUM NEURAL NETWORKS

In network learning algorithms, the weight update between neurons is same as the conventional BP algorithm, which obtains changes and error back propagation by gradient descent method. In addition, to accelerate the training speed and avoid trapping into the local minimum, additional momentum and adaptive learning rate method

are introduced, which enables the avoidance of network's local minimum and rapid convergence. So we have:

$$\begin{cases} w = \{w1_{k,n}, w2_{m,k}, w3_{s,m}\} \\ \Delta w = \{\Delta w1_{k,n}, \Delta w2_{m,k}, \Delta w3_{s,m}\} \end{cases}, \qquad (6)$$

where Δw is the weight's update amount. Then the improved momentum equation of back propagation can be shown as:

$$w(t+1) = w(t) - \Delta w(t), \qquad (7)$$

$$\Delta w(t) = mc \cdot \Delta w(t-1) - (1-mc) \cdot lr \cdot \frac{\partial e}{\partial w}, \qquad (8)$$

where $w(t-1)$ is the weight before training and $w(t+1)$ after training; mc, lr are respectively momentum factor and learning rate. So they can be updated as below:

If $e(t) > e(t-1) \times em$ then:

$$mc = mc_1, \; lr = lr \times d_{1r}. \qquad (9)$$

If $e(t) < e(t-1)$ then:

$$mc = mc_2, \; lr = lr \times i_{1r}, \qquad (10)$$

where em is the maximum error rate; mc_1, mc_2 are momentum factors; d_{1r}, i_{1r} are decrement and increment of learning rate respectively And the update equation of quantum interval is:

$$\langle o_{i,q} \rangle = \frac{1}{|c_q|_{x_k \in C_q}} \sum_{x_k \in C_q} o_{i,k}, \; v_{i,k,s} = o_{i,k,s}(1 - o_{i,k,s}),$$

$$\langle v_{i,q,s} \rangle = \frac{1}{|c_q|_{x_k \in C_q}} \sum_{x_k \in C_q} v_{i,k,s},$$

where, $\Delta\theta_v$ is the update amount of quantum interval; $o_{i,k}$ is the output value of the *i-th* neuron of the first hidden layer when the input vector is x_k; $o_{i,k,s}$ is the output value of the *i-th* neuron of the first hidden layer when the input vector is x_k, the quantum interval as *s*.

3 Evidence theory

Evidence theory, proposed by Dempster in 1967 and developed by Shafer in 1976, is also known as D/S evidence theory. It can be used to deal with the uncertainty caused by the Unknown. With belief function as as a metric, it can build belief function by relaxing probability restriction of event, with no concern for the precise and inaccessible functions. Once the constraint is limited to strict probability, this theory becomes probability theory. Due to its strong ability of handling uncertain information, D/S evidence theory has always been the important means for multi-sensor information fusion. However, its

foundation, BPA (basic probability assignment) is not easily determinable, which constrains the application of Evidence theory. This study proposed a fault diagnosis based on the information confusion with the integration of various intelligence theories and DA evidence theory. The respective evidence body is composed by SVM, grey relational entropy grey relation entropy, and D/S evidence theory is used to complete evidence integration for partial diagnosis of diagnosis module groups. Finally, the decision diagnostic results can be get. The organic combination of such intelligence theory as support vector machine and Evidence theory has improved the comprehensive diagnosis performance and fault diagnostic effects of complex systems.

D/S evidence theory, proposed by Dempster and then developed by his students Shafer, is an extension of the classical probability theory. According to this theory, Collect all the evidences might influence the Assumption, and then divide them into some relative independent Meta evidence (evidence ingredient with single factor), thus forming an evidence space. Next, assign all the possible combinations of meat evidences with a value satisfying certain constraints weaker than probability restriction. Finally, a function defined in power set of evidence space can be obtained, called basic probability assignment function.

Since Subset of evidence space is not independent and the constraint condition is weaker than probability, basic probability assignment function is not probability. In order to obtain reliability similar to probability, another function based on basic probability assignment function is designed, namely the genus probability function. It meets the constraint condition similar to probability, can be used to represent the reliability of the evidence. Meanwhile, the Unknown and Uncertainty can be distinguished by basic probability assignment function, which is more profound and detailed description of Inaccuracy.

Set U as a recognition framework, which contains all the possible results known by people. A is a subset of U. Then function $m: 2 \to [0,1]^U$ meets the following conditions:

1. $m(\phi(0)) = 0$;

2. as $\sum_{A=U}(A) = 1$, A is said to be the basic probability assignment of $m(A)$, and function $Bel: \to [0,1]$, $Bel(A) = \sum_{B \subset A}(B)(\forall A \subset U)$, the belief function on U.

If $m(A) \geq 0$, then A is termed focal element of belief function Bel, the union of all focal elements are called kernel. For combination of evidence groups, Dempster-Shafer has provides the following rules. Set m_1, m_2 as two basic probability assignments defined in the same recognition framework U, Bel_1 and Bel_2 are the belief functions in U, and the focal elements are $A_1, \ldots, A_k$ and $B_1, \ldots, B_r$ and their combinations can get a new basic belief assignment [11-16],

when $\quad A_i \cap B_j = \phi$, $\qquad K = \sum_{i,j} m_1(A_i) m_2(B_j)$,

$$M(P) = \sum_{i,j} m_1(A_i) m_2(B_j), \quad \forall A_i \cap B_j = P, \quad P \neq \phi, \quad \text{when}$$

$P = \phi$, $m(P) = 0$

Taking into account the mentioned above fusion equation, separate evidence from different sources can be combined to obtain more accurate information.

Plausibility function: the plausibility function of proposition A is: $PL(A) = 1 - Bel(\overline{A}) = \sum_{A \cap B = \phi} (B)$.

Plausibility function is also called upper limit function, which represents uncertainty measure for trust degree of proposition. Belief function and plausibility function constitute belief interval of proposition A.

4 Combination forecasting based on quantum neural networks and evidence theory

To compensate for the shortcomings of single diagnostic method, this study combines quantum neural networks with information fusion to make a more accurate comprehensive diagnosis.

For combination of quantum neural networks and information fusion, the key point is to take the output value of quantum neural networks as an evidence, and then make a comprehensive diagnosis by DS evidence theory. The specific implementation method is: firstly, quantum neural networks fusion is processed to get the preliminary fusion results; then normalization for these results are carried out; next take the output value as basic probability assignment for proposition in D/S theory recognition framework; finally, apply the make a comprehensive diagnosis by combination rules. Figure 2 shows repeated fault diagnosis mode for transformer, which is also the overall research program structure of this study.

Diagnostic results

↑

Comprehensive Diagnosis

↑   Secondary fusion

Fault Diagnosis

↑   D/S evidence theory

Secondary DNN diagnosis

↑

Primary fusion

↑   D/S evidence theory

Primary DNN diagnosis   Primary fusion

↑

Data collection

↑
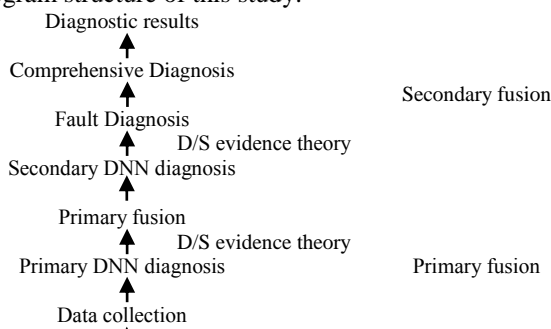
FIGURE 2 Repeated fault diagnosis mode for transformer

This is a secondary fusion structure. On the basis of the primary fusion structure which makes transformer fault diagnosis with quantum neural networks and Evidence theory, this secondary structure makes repeated fault diagnosis with the two fusion algorithms.

In the primary fault diagnosis, the input information combining D/S theory and quantum neural networks is concentration of H_2, CH_4, C_2H_6, C_2H_4 C_2H_2, expressed in 5 neurons. There are two kinds of input forms: input directly various gas concentrations or normalize samples. While the former easily leads to excessive large sample space, which in turn, leads to oversized network scales, affecting network's normal training and diagnosis. So this study selects the latter form, and takes the ratio of gas concentration to the sum of 5 gas concentrations as the input information. The optimal number of the hidden layer nodes is sought on actual situation. There are 6 neurons in the output layer, representing 6 fault types that networks can diagnosis, i.e. Low-temperature overheat, Middle-temperature over, High-temperature overheat, partial discharge, spark discharge and arc discharge.

Based on the primary fusion, the input information of second fault location diagnosis combining D/S theory and quantum neural networks is respectively: low-temperature overheat, middle-temperature over, high-temperature overheat, partial discharge, spark discharge and arc discharge; the optimal number of the hidden layer nodes is sought on actual situation.

5 Experimental study

The encoding and decoding problem is one of the basic problems of neural network training. In this study, a 10-23-10 codec is trained to test the convergence properties of improved quantum neural network mode. There are 10 different input modes, each mode with only one bit as 0 and the rest as 1. The input mode is required to be the network's output value. Set the target literacy accuracy as 0.001, replacement learning rate of weight and threshold value as 0.9, replacement learning rate of quantum interval as 0.7. The initial weight and threshold value are selected randomly. Below is the convergence curve (Figure 3).
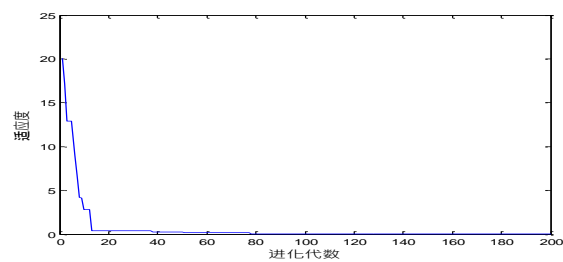
FIGURE3 Convergence curve of quantum neural networks.

As it is seen in Figures 4 and 5, 200 groups of data are selected as samples, among which 33 are test data. After network training, the training data is put into the trained quantum neural networks as input values. As a result, only two samples are misjudged. Moreover, when verified with training samples, also only two samples - two types of transformer fault - are misjudged, a diagnosis precision of 93.7%. The resulting quantum neural network can be used for more sample forecasts.

To select fault features of neural network modes, it is required the maximal fault information be concluded in fault feature samples. Therefore, depth analysis of fault mechanism and fault information transfer relationship is needed. Furthermore, select the best indication of fault

characteristics and dismiss irrelevant ones, to ensure the generation of the smallest quantum neural networks.
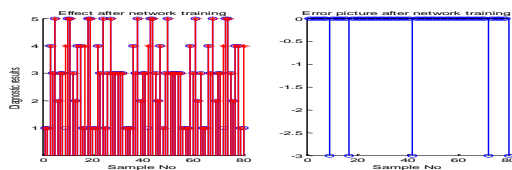
FIGURE 4 Transformer fault classification effect based on quantum neural networks (training sample data)
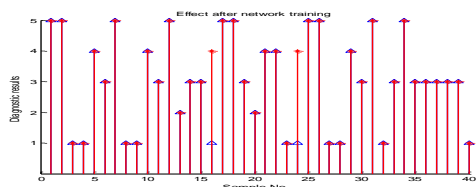
FIGURE 5 Transformer fault classification effect based on quantum neural networks(test data)

In this study, for the same fault, BP neural network is applied to analyse transformer fault diagnosis on the same samples. It includes tree network structures: the input lay, the hidden layer and the output layer. 5 input nodes and 6 output nodes are used based on original data.

After repeated training with BP network gradient descent algorithm of BP network, the result shows that 17 is the optimal number of nodes for the hidden layer, with faster convergence and smaller network errors.

5 Conclusions

By using quantum computation theory, quantum neural networks have more parallel processing power and larger storage capacity than ANN. Quantum theory is the outcome of classical physics to the micro-level. Quantum system is the microscopic system basis of all the physical process, also the basis of biological and psychological ones, which has the similar dynamic characteristics as biological neural networks. By combing quantum theory with ANN, Quantum neural networks can better simulate and interpret human brain's information process, which is a quantum extension and evolution of ANN.

Acknowledgements

References

[1] Cao D 2005 Gas Analysis Diagnostics and Troubleshooting in Transformer Oil *Beijing: China Electric Power Press* 2005 *(in Chinese)*
[2] Zhou D, Ye Y 2000 Modern Fault Diagnosis and Fault Tolerant Control *Beijing: Tsinghua University Press (in Chinese)*
[3] Wang Dg, Zhang Y 2004 Intelligent Fault Diagnosis System *Beijing: Metallurgical Industry Press (in Chinese)*
[4] Jiao L 2005 Application and Implementation of Neural Networks *Xi'An: Xidian University Press (in Chinese)*
[5] Zheng R 2007 Fault Diagnosis Technology of Power Transformer Based on Grey System Theory *Jilin: Jilin University (in Chinese)*
[6] Gao Y 2007 Power Transformer Fault Diagnosis and Life Prediction Method *Xi'An: Xidian University (in Chinese)*
[7] Yin T 2006 Research on Transformer Fault Diagnosis Based on Grey Neural Network *Electrotechnical Application* **25**(10) 15-17 *(in Chinese)*
[8] Ventura D, Martinez T 1997 An artificial neuron with quantum mechanical properties *Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms* 482-5
[9] Behrman E C, Steck J E, Skinner S R 1999 A spatial quantum neural computer *Proceedings of the International Joint Conference on Neural Networks* Washington D.C.

Authors

Qiang Song, Xintai City, Shandong Province, China

Current position, grades: associate professor at Anyang Institute of Technology.
University studies: Master's degree at Kunming University of science and technology in 2006.
Scientific interests: intelligence control theory and rough set theory.

Ai-min Wang, born in 1957

Current position, grades: professor at Anyang normal college.
University studies: Doctor's degree at Wuhan institute of Technology.
Scientific interests: SVM and rough set of intelligent decision-making research.

Survey of research directions in fuzzy cognitive map

Nan Ma[1]*, Yun Zhai[2], Bingru Yang[3]

[1]*College of Information Technology, Beijing Union University, 100101, Beijing, China*

[2]*E-Government Research Center, Chinese Academy of Governance, Beijing, 100089, China*

[3]*School of Computer and Communication Engineering,University of Science and Technology Beijing, Beijing 100083, China*

Abstract

Fuzzy Cognitive Map as a generally recognized intelligent tool has been widely used in the data mining and the machine leaning fields. This paper reviewed the present research situation of Fuzzy Cognitive Map at home and abroad in recent years. Firstly, it reviewed the development progress of Fuzzy Cognitive Map with its advantage over other fuzzy learning methods, followed by the research papers indexed by EI database, ACM database and the SCI database. Then it summarized the classification and the learning methods of Fuzzy Cognitive Map. Furthermore, the applications of Fuzzy Cognitive Map are analysed finally. We hope our work help the people to have a general understanding and push the Fuzzy Cognitive Map forward.

Keywords: fuzzy cognitive map, development, classification, learning methods

1 Introduction

1.1 DEVELOPMENT PROGRESS

In 1976, Axelord proposed a three-valued cognitive map [1]. It consists of two different arcs, one positive, and one negative. The positive indicates the causation have a same changing direction, negative contrast, and the arcs represent the relationship between the different concept nodes. Concept nodes specified action, causes, consequences, purpose, feelings, tendencies and trends of the system, which reflect the properties, performance, quality and status of the system. The relationship of the concept nodes indicates the relevance and effect of them, which expressed as arcs with arrows. In addition to the positive and negative, the strength of the relevance is expressed in numerical, who we called the weight of arc [2]. It can be applied to decision-making model of the application system.

In 1986, Kosko and others, who on the basis of Axelord, combining the fuzzy set theory, proposed a fuzzy cognitive map (FCM) theory, which extends the three-valued logic relationship between nodes to fuzzy relationship on the interval [-1,1] [3]. This theory carries more information than Axelord's three-valued logic cognitive maps , and points out that the limited input state of FCM can open up a path in the virtual space , and simple FCM path may terminate at a fixed point or limit cycle , and in the complex feedback FCM this path may be terminated in the " chaos " strange attractors [4] .

In 1992, Hagiwara, on the basis of FCM, proposed the extended fuzzy cognitive map (EFCM) [5]. This cognitive map extends the linear relationship between concept nodes to the non-linear relationship, introduced the delay of time

and causation and the weight of conditions. In 1994, Wellman proposed a qualitative probabilistic network [6], which regards the cognitive map as a network with an unknown probability, but this cannot quantify the change degree of causation between concept nodes. In 1999, Carvalho and others proposed a rule-based cognitive map model [7], which can solve the problem of non-monotonic reasoning and non-causal representation. But this model has too many rules, cannot adapt to changes in the external environment, and also has a great amount of computation. In 2001, Yuan and others proposed a dynamic cognitive network [8], which is a continuation of the FCM and has a strong ability to adapt to the environment. In 2003, Luo Xiangfeng, who discussed the probabilistic fuzzy cognitive map [9, 10], by introducing fuzzy measure to the traditional cognitive map model to quantify the impact of causal relationship between concept nodes, a probabilistic model of cognitive maps.

1.2 BASIC CONCEPTS OF FCM

For two different nodes c_i and c_j in the FCM, if the value x_i of the node c_i changes, the value x_j of the node c_j changes consequently, we say that the nodes c_i and c_j have the causality relationship. And we call the arc from the node c_i to c_j a directed arc. Then the node c_i is call the reason node, the node c_i is call the result node. A basic model of FCM is demonstrated in the Figure 1.

Corresponding author e-mail: xxtmanan@buu.edu.cn

Operation Research and Decision Making
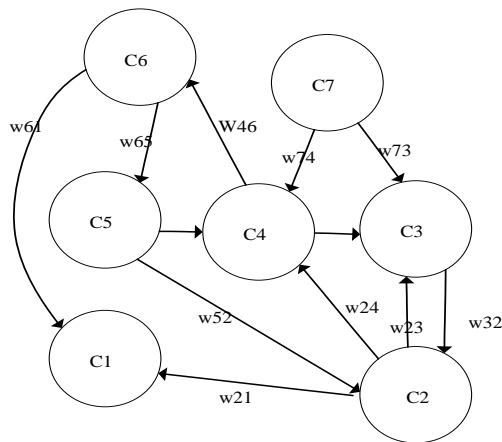
FIGURE 1 Basic model of FCM

Let $C = \{c_1, c_2, ..., c_N\}$ be a finite set of vertices in FCM, where N is the number of nodes, for two any nodes c_i and c_j and the finite set:

$$E = \{e_{11}, e_{12}, ..., e_{1N}, e_{21}, e_{22}, ..., e_{2N}, ..., e_{N1}, e_{N2}, ..., e_{NN}\},$$

in the FCM, each arc has a corresponding weight w_{ij} indicating the influence of node c_i to c_j as shown in Equation (1):

$$W = \begin{bmatrix} w_{11} & w_{12} & ... & w_{1N} \\ w_{21} & w_{22} & ... & w_{2N} \\ ... & ... & ... & ... \\ w_{N1} & w_{N2} & ... & w_{NN} \end{bmatrix}. \tag{1}$$

For any nodes c_i and c_j in FCM, if the two nodes exist a directed relation, the interval [-1,1] can be used to describe the influence degree, i.e., $w_{ij} \in E$ and w_{ij} is the weight of c_i to c_j.

1.3 ADVANTAGES OF FCM

Compared with the neural network model, the advantages of fuzzy cognitive map model is as following:

(i) Each node and arc in fuzzy cognitive map model has a very robust semantics, so that the entire diagram showing a robust semantic, and the results of reasoning is easy to understand. Neural network is a numerical framework, cannot directly represent the structure of knowledge.

(ii) Fuzzy cognitive map model can apply expert knowledge to compensate for the lack of learning data to a certain extent. Because its reasoning is based on matrix operations, so it corresponds to the development of AI that intelligent behaviour driven by data. Neural network does not use expert knowledge of the system itself, and requires a lot of training data, but for those are not easy to get the training data, neural networks will be very limited.

(iii) Fuzzy cognitive map model cannot only represent a semantic network, but also can handle the distribution of knowledge. For any number of knowledge sources it can construct their cognitive maps and can be superimposed on each other arbitrarily, resulting in an overall knowledge

sources joint distribution of knowledge. Neural network does not have the overlay, and it has poor learning ability for system.

(iv) The large-scale fuzzy cognitive map cannot only be utilized to predict, but also can be used to the sensitivity analysis, causal relationship, strategic planning of factor changes. Neural networks are black-box model; the whole system cannot achieve structural analysis.

Compared with traditional methods of knowledge representation and reasoning based on first-order logic, FCM has the following characteristics:

(i) FCM is very convenient to establish, intuitive performance issues, and is suitable for knowledge engineers to interact with experts in the field. Related concepts are connected by arcs together, and the facts, features, and relationships associated with concepts can be deduced through the arc, easy to achieve the explanation of system by way of association.

(ii) FCM can be deduced by digital matrix calculations. Owing to its derivation process is digital, it has more flexibility in terms of calculation and the derivation, the computer can give full play to the advantages of numerical computing.

(iii) FCM supports feedback mechanism, so the result may be a limit cycle, unlike traditional expert system that will only come to a conclusion constant. This makes FCM's ability of knowledge representation and reasoning expanded over traditional expert systems.

Compared with HMM (Hidden Markov Model, HMMs), the differences between them are as following.

(i) Each node in the HMMs represents a specific system state, and therefore, this technique only applies to simulate a finite state system.

(ii) The transfers between states in HMMs is controlled by a set of probabilities, while in FCM the next state is calculated by the conversion function.

1.4 RESEARCH RESULTS

By searching the worldwide ACM, SCI and EI three authoritative database, we found that, from 2004 to 2013, more and more scholars and experts conducted in-depth study in fuzzy cognitive map and achieved fruitful results. Specifically is as showed in Figure 2-4:
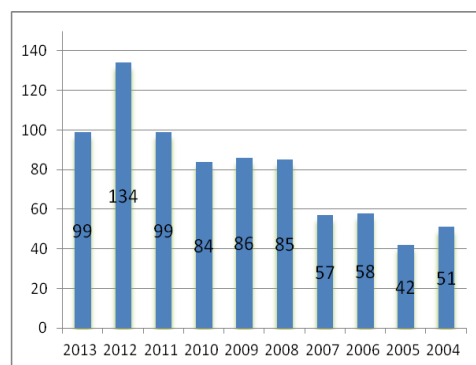
FIGURE 2 EI Database
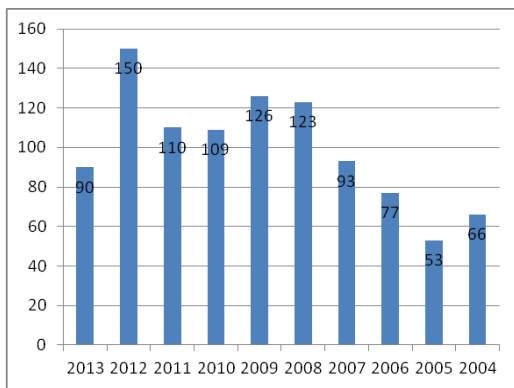
FIGURE 3 SCI Database
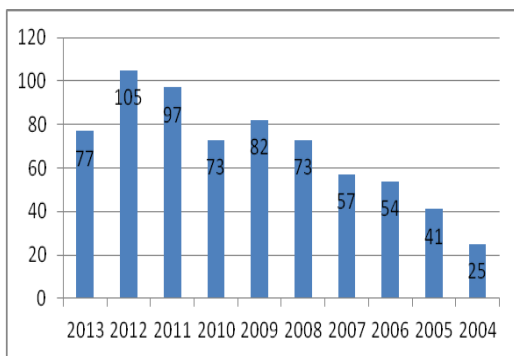
FIGURE 4 ACM Database

By searching the worldwide ACM, SCI and EI three authoritative database, we found that, from 2004 to 2013, more and more scholars and experts conducted in-depth study in fuzzy cognitive map and achieved fruitful results. Specifically is as showed above.

1.5 APPLICATION AREA

As can be seen from the development of FCM, researches in fuzzy cognitive map constantly goes deeper. It is also expanding its application areas, for example Cole and others proposed how to apply FCM to teaching and research [11]. Stylios and others proposed the FCM is applied to supervisory control system model [12]. Lee and others applied FCM to the web mining [13]. Min and others applied the concept of FCM to intelligent decision support [14]. Kardaras and others applied the FCM for commercial modelling [15].

2 Classification of fuzzy cognitive map

2.1 PROBABILISTIC FUZZY COGNITIVE MAP

Probabilistic Fuzzy Cognitive Map (PFCM) was first proposed by the Luo XiangFeng and others. PFCM introduced conditional probability measure in a causal relationship between concept node, the model can represent a qualitative and fuzzy causal relationship between concept nodes, but also represent conditional probability causation between concept nodes, and can degenerate into fuzzy cognitive map.
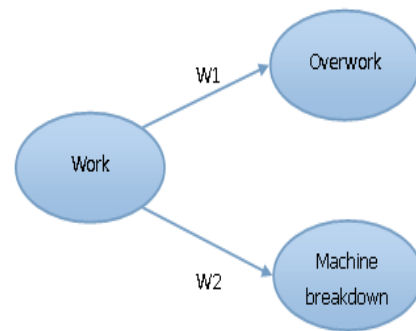
FIGURE 5 Causality of concept nodes

As shown in Figure 5, it is uncertain that the measure of causation between "work" and "overworked" or "machine breakdown" concept nodes. w_1, w_2 measure is not only associated with "labor" concept node's own state values (intensity of labor and labor time). While w_1 is also associated with the values of "worker's body" concept node's state (health, weakness), w_2 is also associated with the values of concept nodes "quality of the machine" (or inferior quality, texture condition) and "machine" texture condition (manufacturing materials).

As FCM cannot represent the dynamic dependencies of concept node's status value on the measure of causation between concept nodes, PFCM first introduced the conditional probabilistic relationship in FCM. In order to introduce the probabilistic relationship to FCM, Stylios proposed the following formula to solve the dependencies of probabilistic between concept nodes in cognitive map feedback cycle, which form a FCM model with time and memory status:

$$Vc_j(t) = f\left(\sum_{\substack{i=1 \\ i \neq j}}^{N} Vc(t)w_{ij} + rVc_j(t-1)\right), \qquad (2)$$

where the $Vc_j(t)$ of the formula c_j node means the state value at time t; γ works as a factor of the last time.

Reference [12] renewed the previous formula with the probability of getting a memory function and the dynamic characteristics of fuzzy cognitive map model:

$$Vc_j(t) =$$

$$f\left(\sum_{\substack{i=1 \\ i \neq j}}^{N} Vc(t)P(w_{ij}(t)|Vc_i(t), Vc_m(t), ..., Vc_n(t)) + rVc_j(t-1)\right). \quad (3)$$

2.2 RANDOM FUZZY COGNITIVE MAP

In 1943 psychologist W. Mcculloch and mathematical logician W. Pittsin on the basis of analysis and summary of the basic characteristics of neurons, first proposed a mathematical model of neuron. This model is still in use, and directly impacted the research progress in this area. Artificial neural networks have a preliminary adaptive and self-organizing ability. Changing the synaptic weights in the learning or training process in order to meet the

279

demands of the surrounding environment. The same network can have different functions because of different learning ways and content. Neural network provides a good tool to Fuzzy Cognitive Map.

References [16, 17] proposed a Random Fuzzy Cognitive Map (RFCM) based on random neural network, and demonstrate its application in process modeling. Based on the activation probability of network neurons, this model achieves reasoning process by numerical calculation instead of symbolic derivation. In RFCM, the arc defines the relationship between the concept nodes, and describes the causal process. The experiment demonstrated that the random fuzzy cognitive map and early fuzzy cognitive map produced approximation results, and less iteration. On this basis, the reference [18] proposed a RFCM based Adaptive Random Fuzzy Cognitive Map (ARFCM). ARFCM transform causal ambiguity into a causal relationship between the network mode conversion, with the use of experts to enhance the causal knowledge. Meanwhile, the document also describes how to use ARFCM interpretation model constituted by a dynamic process.

2.3 RULE-BASED FUZZY COGNITIVE MAP

Reference [19] proposed a Rule-Based Fuzzy Cognitive Maps (RBFCM). RBFCM, based on the traditional fuzzy network with feedback, can be used for modeling describe real-world dynamic system with feedback chain. Thus, proposed RBFCM makes FCM label cognition more completely. Based RBFCM, reference [9] regards time as an important qualitative entity in cognitive map, that is, implicit represent time in RBFCM through an association program. Time in a dynamic system is undoubtedly crucial, which also reveals an important role of RBFCM. As an extension of RBFCM, reference [18] raised a number of questions about stability around the qualitative dynamic system of real-world modeling. Meanwhile, it proposed the inherent stability as a required attributes concept node in the qualitative system. Reference [19] discussed how to achieve RBFCM causality itself. As the traditional fuzzy operation is defined by causal mapping, and therefore a causal relationship cannot be achieved, which is the literature [20] the main basis. The literature use specific methods to achieve fuzzy causation. This approach allows not both add, delete the concept of nodes and links between the nodes of the concept, but also a great degree of flexibility with a novel fuzzy operation, which can mimic causation "accumulate" attribute, namely Fuzzy Carry Accumulation (FCA).

2.4 COMPLEX FUZZY COGNITIVE MAP

Because of the large size and complexity of the internal links of complex systems to study the "big" FCM become one of the most arduous of the most pressing research.

Reference [21] proposed a structure thought polymerization fuzzy cognitive map of the complex systems are often broken down into subsystems, each subsystem are represented by fuzzy cognitive map, then put each fuzzy cognitive map by aggregating them part of the same get the whole complex system of fuzzy cognitive map, also known as polymerization fuzzy cognitive map. The easiest way is to put the various subsystems of polymerization of the corresponding concept of a causal relationship between nodes by simple arithmetic sum and then make vague standardization. But the weight of each subsystem are not equal weight under normal circumstances, and even may have a different symbol, then how to construct an adjacency matrix has the consistency of a complex system, so that the system can achieve mutual cooperation and coordination is a very difficult thing. Reference [22] proposed a method by means of expert intervention. Mutual cooperation and coordination subsystem is affected by the right weight to the role of concepts, therefore, generally considered by summing the weighting factor, preferably the right or to the importance of trust for Subsystem weight coefficient.

If complex systems concepts or things are hierarchical, can be layered deal, each level corresponds to a FCM. FCM simple hierarchy has the following features: All nodes are hierarchical concept, but it results in each node and its lower-level node, the node is the reason for its parent node. FCM simple hierarchical structure is more like a tree, it stressed that it would have a close association between the nodes in the same layer, produced only a causal association between the different layers and parent nodes and child nodes, there is no causal relationship between grandparent and grandchild nodes. Reference [23] with FCM construct integrated two-stage hierarchy process modeling for complex radiotherapy proposed algorithm and the interaction between the two levels of monitoring the implementation of decision support , but did not begin a general study of the levels of FCM, only to specific examples of doing some research. In reference [24], the cognitive map model is applied to the field of corporate credit risk assessment established based on statistical learning and expert knowledge combined with multilayer fuzzy cognitive map. Model of multi- fuzzy cognitive map and its application had a very good try, but their discussion multilayer fuzzy cognitive map is to require anti- reflexive, symmetric and recovery, therefore, has the particularity.

Literature [25] put forward the theory of complex systems and packet decomposition algorithm FCM. That is the first node of the original group cognitive map, and then on the set of nodes construct fuzzy cognitive map providers. Such a decomposition of an original supplier FCM and several meaningful sub FCM. FCM on the original analysis of the cognitive map into the analysis and sub-suppliers of FCM. Causal reasoning quotient FCM provides overall information on the original.

3 Learning methods of fuzzy cognitive map

There are two constructor FCM: artificial construct and calculate structure. Artificially constructed using expert knowledge and experience to build applications FCM model; computing structure is to learn from the historical data to automatically establish FCM model.

Artificial construct ways of modelling process consists of three steps: identify the problem domain key concepts; identify the presence or absence of a causal relationship between these concepts; estimated value of relationships. Because of the way too dependent on artificial construct subjective consciousness and expert domain knowledge makes it difficult to build. Thus, in order to produce a calculation based FCM learning methods.

Currently, the two main types of computing structure FCM branches: one is based on Hebbian learning styles; second is based on the theory of evolution of learning. As shown in Table 1, the former including DHL (Differential Hebbian Learning), BDA (Balanced Differential

Algorithm) NHL (Nonlinear Hebbian Learning), DD-NHL (Data-Driven Nonlinear Hebbian Learning) and AHL (Active Hebbian Learning). The latter has GS (Genetic Strategy), PSO (Particle Swarm Optimization), SA (Simulated Annealing), RCGA (Real Coded Genetic Algorithm).

In addition to these methods are selected BDA continuous conversion function; except GS initial vector is obtained, other learning algorithms are FCM link target matrix; PSO method needs only a plurality of GS and the state vector, and this is not available in some areas. Based Hebbian algorithm uses an unsupervised learning mechanism, based on the weight associated with a single sample of the initial system state data FCM's training so as to reach a steady state (fixed-point state). Different algorithms that adjust each side of the weights way, because it does not require a lot of computing training faster, but they are simple systems for, FCM produced only node includes several concepts.

TABLE 1 Comparison of FCM learning methods.

Algorithm	Learning goal	Human intervention	Type of data used	Transformation function	Learning type
DHL	Connection matrix	No	Single	N/A	Hebbian
BDA	Connection matrix	No	Single	Binary	Modified Hebbian
NHL	Connection matrix	Yes&No	Single	Continuous	Modified Hebbian
DD-NHL	Connection matrix	Yes&No	Single	Continuous	Modified Hebbian
AHL	Connection matrix	Yes&No	Single	Continuous	Modified Hebbian
GS	Initial vector	No	Multiple	Continuous	Genetic
PSO	Connection matrix	No	Multiple	Continuous	Swarm
SA	Connection matrix	No	Single	Continuous	Simulated annealing
RCGA	Connection matrix	No	Single	Continuous	Genetic

4 The applications of fuzzy cognitive map

The existing learning algorithms based on FCM in data mining are mainly used in classification, where each attribute of the target is regarded as a node in the FCM, the initial weight of every edge is given by experts, and experts need to clear the input node and the output node. A typical classification model is as demonstrated in the Figure 6.
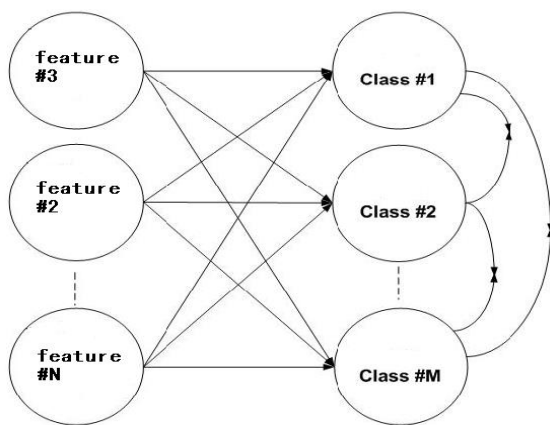
FIGURE 6 A typical classification model

The learning algorithm for fuzzy cognitive map is based on the genetic algorithm with real coding method, such that this learning algorithm tend to over-fit with

different initial values. Therefore, they implemented two criteria to reduce the over - fitting probability, i.e., the average error of one input value in all cycles and the average error of different input values in all cycles. The former calculation method is computed using Equation (4) and the latter using Equation (5):

$$error\_behavior =$$

$$\frac{1}{P(K-1)N} \sum_{p=1}^{P} \sum_{t=1}^{K-1} \sum_{n=1}^{N} \left| A_n^{act}(t) - A_n^{des}(t) \right|, \tag{4}$$

$$error\_initial = \frac{1}{(K-1)N} \sum_{t=1}^{K-1} \sum_{n=1}^{N} \left| A_n^{act}(t) - A_n^{des}(t) \right|. \tag{5}$$

The references [28, 29] realized a FCM-GT based on the bladder tumour classification, the weighted matrix of FCM-GT is as shown in Figure 7.
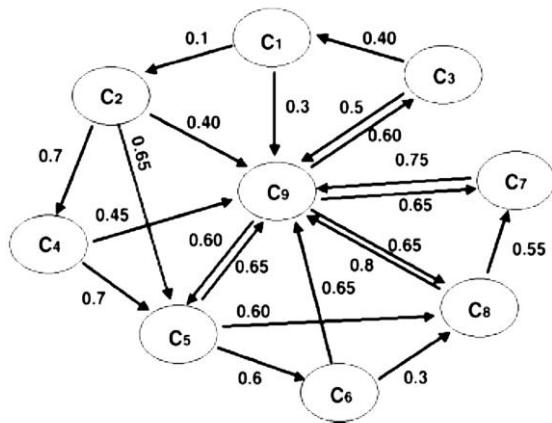
FIGURE 7 The weighted matrix of FCM-GT

5 Conclusions

As an important method of soft computing, the research of FCM has attracted more and more attentions, and just as analysed in this paper, the study of FCM has made great progress. In this field, however, there are still many problems to be solved, the focus will focus on the following aspects:

1) The importance of time in FCM. Time scale is crucial for the dynamic system, the issue of node segmentation and polymerization based on the time scale is worthy of more attention to concept of [28].

2) FCM integration strategy. Based on the existing ensemble theory, the ensemble theory is introduced into the FCM integrated framework in the progress of FCM.

3) To expand the application field of FCM. FCM has been applied to some simple environment, how to put it into more complex environment such as the virtual world with many experts need researchers pay more attentions, when the FCM perhaps becomes a powerful tool for intelligent decision support.

4) So far, the research of FCM is limited to the algorithms, but the theoretical basis is still lack and the research fruits are not enough rich to people, most research is on the basis of the experiment method lack of mature theoretical support, so it is very important to study further.

5) The mechanism, model theory of FCM may produce a breakthrough point, i.e., FCM will integrated with the nonlinear dynamics, streaming media dynamics, multi relational mining, manifold learning and directed hypergraph model.

Acknowledgments

References

[1] Axelrod R 1976 Structure of Decision:the Cognitive Maps of Political Elites Princeton University Press New Jersey
[2] Fantacci R, Nannicini S 2000 *IEEE Journal of Selected Areas in Communications* **18**(8) 1441-54
[3] Kosko B 1986 International Journal of Man-Machine Studies 24 65-75
[4] Chun M L 2006 Model method and application *Study of Fuzzy Cognitive Map Donghua University: Shanghai Chapter* 2 (*in Chinese*)
[5] Hagiwara M 1992 Proceedings of the International Conference on Fuzzy Systems *IEEE Press San Diego* 795-801
[6] Marchant T 1999 *European Journal of Operational Research* **114**(3) 626-37
[7] Carvalho J P, Tome J A 1999 Evolutionary Computation & Fuzzy Logic for Intelligent Control *Knowledge Acquisition & Information Retrieval IOS Press USA* 276-81
[8] Carvalho J P, Tome J A 2000 *Proceedings of the International Conference of the North American Fuzzy Information Processing Society Atlanta IOS Press USA* 407-11
[9] Feng L X, Jun GAO 2003 *Journal of Computer Research and Development* **40**(7) 925-33 (*in Chinese*)
[10] Feng L X, Gao J 2003 *Journal of University of Science and Technology of China* **33**(1) 26-33 (*in Chinese*)
[11] Cole R J, Persichitte K A 2000 *International Journal of Intelligent Systems* 15 1-25
[12] Stylios C D, Groumpos P P 1999 Computers in Industry **39**(3) 229-38
[13] Lee K C, Kim J S, Chung N H, Kwon S J 2002 *Expert Systems with Application* **22**(3) 197-211
[14] Min, H Q, Hui J X, Lu Y S, Jia Z J 2006 *Proceedings of the International Conference on Intelligent Agent Technology Hong Kong, China IOS Press USA* 321-5
[15] Kardaras D, Mentzas G 1997 Advances in Industrial Engineering Applications and Practice II 63-68
[16] Aguilar J 2001 *Lecture Notes in Artificial Intelligence* 2070 333-8
[17] Aguilar J 2001 *Proceedings of the International Conference of the International Joint Conference on Neural Networks Washington IOS Press USA* 1380-5
[18] Aguilar J 2003 *International Journal of Computational Cognition* **1**(4) 91-107
[19] Groumpos P P, Stylios C D 2000 *Chaos Solitons and Fractals* **11**(1-3) 329-36
[20] Stylios C D, Groumpos P P 1999 *Journal of Intelligent and Robotic Systems* **26**(3-4) 389-403
[21] Ndousse T D, Okuda T 1996 *Proceedings of the IEEE International Conference on Communications* 1558-1562
[22] Feng L X 2003 *Dissertation Submitted to Hefei University of Technology for the Degree of PhD* 1-67 (*in Chinese*)
[23] Elpiniki I P, Stylios C D, Groumpos P P 2003 *IEEE Transactions on Biomedical Engineering* **50**(12) 326-39
[24] Stach W, Pedrycz W, Kurgan L 2004 *NAFIPS* 28-33
[25] Yun Z G, Rong M X, Ru Y B 2007 *Computer Science in China* **34**(4) 129-32 (*in Chinese*)
[26] Papageorgiou E I, Spyridonos P P, Glotsos D Th, Stylios C D 2008 *Applied Soft Computing* 8 820-8
[27] Papageorgiou E I, Georgoulas G, Stylios C D, Nikiforidis G N, Groumpos P P 2006 *Proceedings of the International Conference of International Conference on Knowledge-Based & Intelligent Information & Engineering Systems* 515-23
[28] Wojciech F, Alicja W D 2009 *Proceedings of the International Conference on Human System Interactions IEEE Press Catania* 13-20

Operation Research and Decision Making

Authors

Nan Ma, born in December, 1978, China

Current position, grades: PhD, Beijing Union University.
Scientific interest: knowledge discovery.
Publications: 5 books and more than 20 papers.

Yun Zhai, born in June, 1979, China

Current position, grades: PhD, Chinese Academy of Governance.
Scientific interests: knowledge discovery.
Publications: more than 30 papers.

Bingru Yang, born in March, 1941, China

Current position, grades: University of Science and Technology Beijing.
Scientific interests: knowledge discovery.
Publications: 10 books and more than 500 papers.

Exploring the relationship between inventory level and bullwhip effect in the supply chain

Lin Li*

School of Management, HeFei University of Technology, Anhui PR China

Anhui finance & trade vocational college, Anhui PR China

Received 1 July 2014, www.cmnt.lv

Abstract

This paper intends to work out an expected inventory level formula for the retailer in the two-stage supply chain. It aims at disclosing the quantitative relationship between bullwhip effect and expected inventory level and does analysis of simulation experiment. The model supposes the market demand faced by the retailer follows the autoregressive process AR(1) and that the retailer makes anticipation of the market demand during the replenishment lead time by mean square error method. Moreover, if the interference factor follows normal distribution with mean 0 and variance σ^2, market demand in every period, demand estimation during the replenishment lead time and the order quantity made by the retailer are all proved to follow normal distribution.

Keywords: bullwhip effect, first-order autoregressive model, supply chain management, inventory level, demand forecasting

1 Introduction

Bullwhip effect in the supply chain is called demand amplification effect [1-3]. When the demand variability increases as one move up the supply chain, the bullwhip effect occurs. Small variations in consumer demand can result in large variations in upstream orders and inventory. This process parallels to the bullwhip that swings much greater at the end than at the beginning. In some industries, bullwhip effect is also called "Forrester Wheel Effect" [4], "Whiplash" or "Whipsaw" effect.

Most of the previous researches on the bullwhip effect in the supply chain focus on demonstrating its existence, identifying its possible causes, quantifying the bullwhip effect and providing measures to reduce such effect. This paper considers a two-stage supply chain consisting of a single manufacturer and a single retailer, it differs from previous research in several ways. First, it supposes that the market demand faced by the retailer follows the autoregressive process AR(1) and the retailer makes anticipation of future market demand by mean square error method. Second, under the premises that the white noise follows normal distribution, market demand in every period, demand estimation of the replenishment lead time and order quantity are all proved to follow normal distribution. Finally, this paper intends to work out an expected inventory level by mathematical analysis and finds out a linear relationship between bullwhip effect and expected inventory level. As a result, it proposes the concept of ideal expected inventory level that serves as an indicator of measuring the optimal structure of the supply chain.

The rest of the paper is organized as follows. The next section provides a brief survey on the related literature. Section 3 describes the supply chain model and gives the main results of the paper. Section 4 presents analysis of experiment simulation. Section 5 concludes the paper and discusses the future research.

2 Literature Review

The literature on the bullwhip effect in the supply chain is extensive, so we only provide a description of major classes of models. Lee et al. [5] study the replenishment system in the supply chain where the bullwhip effect occurs. Enterprises in the downstream transmit information to enterprises in the upstream by making orders, which steers the latter's production and inventory decisions. But the variance of orders is larger than that of the real market demand, causing that the information distortion is amplified from downstream to upstream. This paper accounts four causes of the bullwhip effect: demand forecast updating, rationing and shortage gaming, order batching and price fluctuation. Literature [6] quantifies the bullwhip effect in a simple two-stage supply chain consisting of a single manufacturer and a single retailer. It establishes a model that helps to explain two causes of the bullwhip effect, namely, demand forecasting and order lead times. It also expands these results to multi-stage supply chains with and without centralized customer demand information, and proves that the bullwhip effect can be reduced by centralizing demand information. However, it cannot be completely eliminated.

Cachon et al. [7] set up a two-stage supply chain model in which a single supplier and multiple independent

retailers. In this model, the supplier has limited capacity, and retailers are privately informed of their optimal stocking levels. They show that some allocation mechanisms induce the retailers to place their optimal order and a manipulation mechanism may lead the supplier to build more capacity. Besides that, Kelle and Milne [8] study factors of order batching and examine the effect of (s,S) ordering policy on the order variability in a supply chain, they provide quantitative tools for the estimation of the variability increase. Lambert, Copper and Pagh [9] concentrate on operationalizing the supply chain management framework, they present that managing the supply chain involves three closely interrelated elements. Some other researchers as Ryan, Baganha, Cohen and Graves focus on how to optimize the technology of treating market demand information and information sharing mechanism [10-12]. Jaipuria et al. [13] propose an integrated approach of discrete wavelet transforms analysis and artificial neural network for demand forecasting, this model can improve the forecasting accuracy and be applicable to both linear and non-linear data series. Luong [14] studies the effect of autoregressive coefficients and lead time on bullwhip effect for a two stage supply chain, the retailer employs a base stock policy for inventory management using first order autoregressive model. Duc et al. [15] study the effect of existence of a third-party warehouse on bullwhip effect in a supply chain, and assume that the demand process as first order autoregressive model. For a list of updated results, the reader is referred to [16-18].

3 A Supply Chain Model

As many researches, this paper supposes that the retailer makes the decision of replenishment at the end of each fixed period and gives orders to the manufacturer in the upstream at the same time. The time interval between two neighbouring replenishment is same and replenishment occurs when $t = -\infty,...,-1,0,1,n-1,n,n+1,...,+\infty$. Suppose the replenishment lead time is l, indicating that the retailer will receive commodities from the manufacturer after a delay of l periods. Suppose the market demand faced by the retailer is characterized by the autoregressive process AR(1):

$$D_n = \mu + \rho D_{n-1} + \varepsilon_n, \qquad (1)$$

where D_n refers to the market demand or sales in replenishment period n. μ is a non-negative constant, indicating the average level of the market demand. ρ is a correlation parameter with $|\rho| < 1$, indicating demand of two neighbouring replenishment periods. The closer ρ is to 1, the more correlation the demand of two neighbouring periods has. ε_n is the inference noise in period n and independent from demand D_n in period n. $\varepsilon_n, n = -\infty,\cdots,-1,0,1,\cdots,+\infty$ are independent from each other. ε_n includes all information in period n that cannot

be explained by demand in past periods. In other words, it is only related to market environment. Thus, we think ε_n is independent from historical demand $D_{n-1}, D_{n-2}, D_{n-3},...$

3.1 FORECASTING METHOD AND THE INVENTORY POLICY

The retailer makes anticipation of market demand in future replenishment lead time by mean square error method. If replenishment decision is made at the end of period n, the retailer should make demand forecasting. Suppose the market demand of period $n+i$ is $D_{n+i} (i = 1,...,l+1)$, respectively. Then the market demand during the replenishment lead time is:

$$\sum_{i=1}^{l+1} D_{n+i} = \frac{1}{1-\rho}\left\{ \mu \sum_{i=1}^{l+1}(1-\rho^i) + \rho(1-\rho^{l+1})D_n \right\} + \sum_{i=1}^{l+1}\frac{1-\rho^i}{1-\rho}\varepsilon_{n+l+2-i} \qquad (2)$$

By mean square error method, we suppose the demand estimation is D_n^l, then $D_n^l = E\left(\sum_{i=1}^{l+1} D_{n+i} \mid D_n \right)$, where D_n^l is estimated by the conditional expectation of $\sum_{i=1}^{l+1} D_{n+i}$ based on D_n in period n. σ_n^l denotes the conditional variance of $\sum_{i=1}^{l+1} D_{n+i}$, then $\sigma_n^l = \text{var}\left(\sum_{i=1}^{l+1} D_{n+i} \mid D_n \right)$. The inventory level in period $n+l+1$ decided at the end of period n is:

$$y_n = D_n + z\sqrt{\sigma_n^l},$$

where, z refers to service level factor or safety inventory factor. The bigger z is, the more safety inventory should be. Let q_n be the variance of the orders placed by the retailer to the manufacturer, then the order quantity q_n is denoted by $q_n = y_n - y_{n-1} + D_n$.

3.2 ANALYSIS OF THE PROBABILITY DISTRIBUTION

For inference factor ε_n in AR(1) process, we suppose it follows normal distribution with mean 0 and variance σ^2. We shall find out that market demand D_n in every period, demand estimation D_n^l and the order quantity q_n made by the retailer all follow normal distribution.

First, if we repeatedly employ Equation (1), D_n can be expressed as:

$$D_n = \mu + \rho D_{n-1} + \varepsilon_n = \mu \sum_{i=0}^{+\infty} \rho^i + \sum_{i=0}^{+\infty} \rho^i \varepsilon_{n-i}.$$

As ε_n is independent from each other, and $\varepsilon_n \sim N(0, \sigma^2)$, from the additive property of normal distribution:

$$\sum_{i=0}^{+\infty} \rho^i \varepsilon_{n-i} + \sum_{i=0}^{+\infty} \rho^i \sim N\left(\frac{\mu}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right),$$

so $D_n \sim N\left(\frac{\mu}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right)$.

For demand estimation D_n^l, from Equation (2) and the expression $D_n^l = E\left(\sum_{i=1}^{l+1} D_{n+i} \mid D_n\right)$, by simple calculation we can get:

$$D_n^l = \frac{\mu}{1-\rho}\left\{(l+1) - \sum_{j=1}^{l+1} \rho^j\right\} + \frac{\rho\left(1-\rho^{l+1}\right)}{1-\rho} D_n. \qquad (3)$$

It has been proved that $D_n \sim N(\frac{\mu}{1-\rho}, \frac{\sigma^2}{1-\rho^2})$. Observe Equation (3), there is only one normal random variable D_n and the rest are all constants. So D_n^l is also a normal random variable.

For q_n, after calculation we can get:

$$\sigma_n^l = \text{var}\left(\sum_{i=1}^{l+1} D_{n+i} \mid D_n\right) = \frac{1}{(1-\rho)^2} \sum_{j=1}^{l+1}\left(1-\rho^j\right)^2 \sigma^2.$$

Thus, the fluctuation σ_n^l of market demand during the lead time has no correlation with period n. So substitute $y_n = D_n + z\sqrt{\sigma_n^l}$ to the expression of q_n and get:

$$q_n = y_n - y_{n-1} + D_n = D_{n-1} + \frac{1-\rho^{l+2}}{1-\rho}\varepsilon_n.$$

As ε_n is independent from D_{n-1} and both are normal distribution, so q_n is also normal distribution. Calculate the expectation and variance of q_n:

$$q_n \sim N\left(\frac{\mu}{1-\rho}, \left(\frac{1}{1-\rho^2} + \left(\frac{1-\rho^{l+2}}{1-\rho}\right)^2\right)\sigma^2\right).$$

From the above mentioned, we can see the normal random property of the order quantity q_n by the retailer in every period is derived from that of ε_n.

3.3 EXPECTED INVENTORY LEVEL OF THE RETAILER

Next, we will deduce the expected inventory level by mean square error analysis. As we have already known, q_n is a random variable that is a normal distribution with mean $\frac{\mu}{1-\rho}$ and variance $\left(\frac{1}{1-\rho^2} + \left(\frac{1-\rho^{l+2}}{1-\rho}\right)^2\right)\sigma^2$.

The retailer gives an order q_n at the end of each period and receives commodities after the interval of l periods. Suppose the consumption speed per time is v, and we use expectation $E(v)$ to replace v. Under $E(v)$, there are three states of inventory: just running up, surplus and shortage. To obtain the expected inventory level at the end of each period, we should get the probability distribution function of commodity consumption time T_v with the amount of q_n in every period.

$$T_v = \frac{q_n}{E(v)} = \frac{q_n}{\dfrac{E(D_n)}{T}} = \frac{T(1-\rho)}{\mu} \cdot q_n,$$

where T refers to the length of time in every period. q_n is known to follow normal distribution, T_v is also the same. Calculate the expectation and variance, and we can get:

$$T_v \sim N\left(T, \left(1 + \frac{(1+\rho)\left(1-\rho^{l+2}\right)^2}{1-\rho}\right)\frac{T^2\sigma^2}{\mu^2}\right).$$

Let $f(T_v)$ denotes the probability density function of T_v. We can get the product expression of expected inventory level and the time:

$$\begin{aligned}
IT &= \int_0^T \int_0^{T_v} (T_v - t)\frac{\mu}{(1-\rho)T}\,dt \cdot f(T_v)dT_v + \\
&\quad \int_T^{+\infty} \int_0^T (T - t)\frac{\mu}{(1-\rho)T}\,dt \cdot f(T_v)dT_v = \\
&\quad \frac{\mu}{2(1-\rho)T}\int_0^T (T_v - T)^2 f(T_v)dT_v + \\
&\quad \frac{\mu}{(1-\rho)} \cdot \int_0^T T_v f(T_v)dT_v - \frac{\mu T}{2(1-\rho)}\int_0^T f(T_v)dT_v + \\
&\quad \frac{\mu T}{2(1-\rho)}\int_T^{+\infty} f(T_v)dT_v.
\end{aligned} \qquad (4)$$

In this expression:

$$T_v \sim N\left(T, \left(1 + \frac{(1+\rho)\left(1-\rho^{l+2}\right)^2}{1-\rho}\right)\frac{T^2\sigma^2}{\mu^2}\right).$$

It means that the range of T_v is the entire abscissa axis from $-\infty$ to $+\infty$. But in real situation, the consumption time must be greater than 0. So we deduce that the main range of T_v is from 0 to $+\infty$ and the rest can be overlooked. As $f(T_v)$ is a normal density function, the

image of $f(T_v)$ is symmetrical about $T_v = T$. So we can get:

$$\int_0^T T_v f(T_v) dT_v \approx \frac{1}{2} \int_{-\infty}^{+\infty} T_v f(T_v) dT_v = \frac{1}{2} \mu_{T_v} = \frac{1}{2} T,$$

$$\int_0^T f(T_v) dT_v \approx \int_T^{+\infty} f(T_v) dT_v = \frac{1}{2} \int_{-\infty}^{+\infty} f(T_v) dT_v = \frac{1}{2},$$

$$\int_0^T (T_v - T)^2 f(T_v) dT_v \approx \frac{1}{2} \int_{-\infty}^{+\infty} (T_v - T)^2 f(T_v) dT_v =$$

$$\frac{1}{2} \sigma_{T_v}^2 = \frac{1}{2} \left(1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho} \right) \frac{T^2 \sigma^2}{\mu^2} \quad .$$

Substitute the results to Equation (4) and get the expression for expected inventory level × time:

$$IT = \left\{ [1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho}] \cdot \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)} \right\} T.$$

So the expected inventory level of the retailer in every period is:

$$I = \left[1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho} \right] \cdot \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)}.$$

3.4 ANALYSIS OF THE RELATIONSHIP BETWEEN BULLWHIP EFFECT AND EXPECTED INVENTORY LEVEL

In this subsection, we first give the expression of the bullwhip effect in the two-stage supply chain, where the market demand follows the autoregressive process AR(1) and the retailer uses mean square error method to estimate market demand. Based on the definition of the bullwhip effect, we have:

$$B = \frac{\text{var}(q_n)}{\text{var}(D_n)} = \frac{\left(\frac{1}{1-\rho^2} + \left(\frac{1-\rho^{l+2}}{1-\rho} \right)^2 \right) \sigma^2}{\frac{\sigma^2}{1-\rho^2}} =$$

$$1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho}.$$

In the previous section, we derive the expected inventory level of the retailer in every period:

$$I = \left[1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho} \right] \cdot \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)}.$$

Next we will analyse this expression. Recall the definition of bullwhip effect B. If $B = 1$, then the variance

of order quantity is not amplified, and the expression will turn to be $I^* = \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)}$. This is the ideal expected inventory level. It is the lowest inventory level.

But as $\frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho} > 0$, which means there is

always $I > \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)}$, so we can never reach the ideal expected level. However, this concept is significant in that it helps the retailer to measure if there is space to optimize the structure of the supply chain.

Let us take a look at the first term $\left[1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho} \right] \cdot \frac{\sigma^2}{4\mu(1-\rho)}$. It is in fact the product

of the bullwhip effect B and the fluctuation $\frac{\sigma^2}{4\mu(1-\rho)}$.

Then the expected inventory level can be expressed as:

$$I = B \cdot \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)}, \quad B \geq 1.$$

When the market demand follows first-order autoregressive model AR(1) and the retailer uses mean square error method to forecast the demand during the replenishment lead-time, there is a linear correlation between the expected inventory level and the bullwhip effect. μ, ρ and σ are parameters in the market demand process. The bigger ρ and σ is, the higher the expected inventory level and the greater the bullwhip effect will be, the expected inventory level will be linearly steeper. Actually, this explains why the bullwhip effect can cause a large amount of overstock in enterprise and why reducing the bullwhip effect can reduce the expected inventory level and the inventory cost.

Let's see the quantifying expression $B = \frac{\text{var}(q_n)}{\text{var}(D_n)} = 1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho}$ again. It shows that

the bullwhip effect expands along with the increase of the replenishment lead-time. If the manufacturer in supply chain can reduce this lead-time, he can lower the bullwhip effect and reduce the expected inventory level so as to optimize the supply chain structure. This conclusion is reached in many related literatures. But this paper firstly points out the quantifying model that the retailer can reduce the expected inventory level if the bullwhip effect can be lowered.

4 Experiment Simulation and Results

To verify the reasonability of the model in supply chain, we apply Visual C++ to it and analyse the results based on parameter adjustment validation.

287

Suppose the supply chain consists of an automobile part producer and a professional automobile repair shop. According to sales statistics, the demand of part faced by the repair shop follows the first-order autoregressive model AR (1): $d_{NT} = 100 + \rho d_{(N-1)T} + \varepsilon_{NT}$, $\varepsilon_{NT} \sim N(0, \sigma^2)$. The repair shop uses mean square error method to estimate the future demand, and the replenishment lead-time is 4.

Table 1, Table 2 and Table 3 are the results of expected inventory level of the retailer, ideal expected inventory level and the bullwhip effect level from the expressions:

$$I = B \cdot \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)},$$

$$I^* = \frac{\sigma^2}{4\mu(1-\rho)} + \frac{\mu}{2(1-\rho)} \text{ and } B = 1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho}.$$

The abscissa parameter refers to dynamic adjustment of correlation parameter ρ with the fixed step 0.1 in the AR (1) process. The longitudinal parameter refers to the standard variance σ of white noise with the value of 2, 4, 9, 16, 25, 36, 49, 64.

TABLE 1 Expected inventory level of the retailer

σ \ ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	55.580	62.531	71.469	83.389	100.078	125.116	166.847	250.295	500.517
4	55.654	62.625	71.592	83.554	100.313	125.464	167.388	251.180	502.069
9	56.056	63.133	72.254	84.452	101.582	127.347	170.320	255.974	510.472
16	57.136	64.500	74.038	86.869	105.001	132.417	178.212	268.880	533.097
25	59.414	67.382	77.800	91.964	112.209	143.107	194.853	296.093	580.803
36	63.556	72.624	84.641	101.230	125.317	162.547	225.114	345.578	667.554
49	70.377	81.256	95.905	116.490	146.903	194.561	274.947	427.070	810.414
64	80.839	94.498	113.185	139.897	180.015	243.668	351.388	552.074	1029.552

As is shown in Table 1, with the increase of correlation parameter ρ and white noise standard variance σ, the expected inventory level rises up. In addition, experimental simulation analysis also proves that I is more sensitive to σ than ρ does. In other words, the difference between the expected inventory level and the ideal one is not significant when ρ and σ are at a low level. But this becomes significant when ρ and σ increase.

TABLE 2 Ideal expected inventory level

σ \ ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	55.567	62.513	71.443	83.350	100.020	125.025	166.700	250.050	500.100
4	55.600	62.550	71.486	83.400	100.080	125.100	166.800	250.200	500.400
9	55.781	62.753	71.718	83.671	100.405	125.506	167.342	251.013	502.025
16	56.267	63.300	72.343	84.400	101.280	126.600	168.800	253.200	506.400
25	57.292	64.453	73.661	85.938	103.125	128.906	171.875	257.813	515.625
36	59.156	66.550	76.057	88.733	106.480	133.100	177.467	266.200	532.400
49	62.225	70.003	80.004	93.338	112.005	140.006	186.675	280.013	560.025
64	66.933	75.300	86.057	100.400	120.480	150.600	200.800	301.200	602.400

As is shown in Table 2, with the increase of correlation parameter ρ and white noise standard variance σ, the ideal expected inventory level I^* rises up. I^* is more sensitive to σ^2 than ρ does. Considering Table 1 and Table 2, we can see that the difference between the expected inventory level and the ideal one is not significant with the growth of ρ but significant with the growth of σ, indicating that it is more sensitive to σ.

TABLE 3 Bullwhip effect level

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
B	2.222	2.500	2.854	3.314	3.907	4.636	5.412	5.900	5.171

Bullwhip effect model $B = 1 + \frac{(1+\rho)(1-\rho^{l+2})^2}{1-\rho}$ has no correlation with the fluctuation standard variation σ. So there is one row of data in Table 3. The value of B increases along with ρ. From Table 1, Table 2 and 3, we can see that the expected inventory level will be up with the increase of B, it is characterized by positive linear correlation. This proves that the model in supply chain is reasonable.

5 Conclusions

This paper bases itself on the two-stage supply chain consisting of a single supplier and a single retailer. It supposes that the market demand follows the first-order autoregressive model AR (1) and future market demand forecasting made by the retailer with the mean square error method. It reveals that there is a linear correlation between the expected inventory level and the bullwhip effect, which establishes quantity link between the two. It also proposes

the concept of ideal expected inventory level to measure the optimized degree of the supply chain structure. Moreover, it also analyses how parameters of the market demand process affect the inventor`y level. In future research, we can further study the cases that the market demand follows other demand process or the retailer uses other demand forecasting methods.

References

[1] Lee H L, Admanabhan V, Whang S 1997 The bullwhip effect in supply Chains *Sloan Management review* **38**(3) 93-102

[2] Wilding R 1998 The supply chain complexity triangle: uncertainty generation in the supply chain *International Journal of Physical Distribution & Logistics Management* **2**(8) 599-616

[3] Taylor D H 2000 Demand amplification: has it got us beat? *International Journal of Physical Distribution & Logistics Management* 30(6) 515-33

[4] Forrester J W 1968 Industrial dynamics *Management Science* **14**(7) 398-415

[5] Lee H L, Padmanabhan V, Whang S 1997 Information distortion in a supply chain: the bullwhip effect *Management science* **43**(4) 546-58

[6] Chen F, Drezner Z, Ryan J K, Simchi-Levi D 2000 Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information *Management Science* **46**(3) 436-43

[7] Cachon G P, Lariviere M A 1999 Capacity choice and allocation: strategic behavior and supply chain performance *Management Science* **45**(8) 1091-1108

[8] Kelle P, Milne A 1999 The effect of (s, S) ordering policy on the supply chain *International Journal of Production Economics* **59**(1) 113-22

[9] Lambert D M, Cooper M C, Pagh J D 1998 Supply chain management: implementation issues and research opportunities *International Journal of Logistics Management* **9**(2) 1-20

[10] Ryan J K 1998 Analysis of inventory models with limited demand information *Ph.D Dissertation Department of Industrial Engineering and Management Science Northwestern University* Evanston IL USA

[11] Baganha M P, Cohen M A 1998 The stabilizing effect of inventory in supply chains *Operations Research* **46**(3-supplement) S72-S83

[12] Graves S C 1999 A single-item inventory model for a nonstationary demand process *Manufacturing & Service Operations Management* **1**(1) 50-61

[13] Jaipuria S, Mahapatra S S 2014 An improved demand forecasting method to reduce bullwhip effect in supply chains *Expert Systems with Applications* **41**(5) 2395-2408

[14] Luong H T 2007 Measure of bullwhip effect in supply chains with autoregressive demand process *European Journal of Operational Research* **180**(3) 1086-97

[15] Duc T T, Luong H T, Kim Y D 2010 Effect of the third-party warehouse on bullwhip effect and inventory cost in supply chain *International Journal of Production Economics* **124**(2) 395-407

[16] Li Q, Disney S M, Gaalman G 2014 Avoiding the bullwhip effect using damped trend forecasting and the order-up-to replenishment policy *International Journal of Production Economics* **149**(3) 3-16

[17] Li C 2013 Controlling the bullwhip effect in a supply chain system with constrained information flows *Applied Mathematical Modelling* **37**(4) 1897-1909

[18] Cigolini R, Pero M, Rossi T, Sianesi A 2014 Linking supply chain configuration to supply chain performance: A discrete event simulation model *Simulation Modelling Practice and Theory* 40(1) 1-11

Author

Lin Li, born in March, 1967, Hefei, Anhui, China

Current position, grades: Doctoral student at the School of Management, Hefei University of Technology. Lecturer in Anhui finance &trade vocational college, Hefei Anhui, China.
University studies: B.S. degree from Hefei University of Technology, Hefei, China in 1996.
Scientific interest: cloud computing, capital operation and management.

Operation Research and Decision Making

A research into location routing problem based on hybrid genetic simulated annealing algorithm

Chengduan Wang*

School of computer engineering, Weifang University, 261061, Weifang, China

Received 1 July 2014, www.cmnt.lv

Abstract

With the diversified and personalized commodity requirements as well as small batch dispatch and frequent dispatch features under the circumstance of E-commerce, the environment for logistics dispatching system becomes increasingly complicated and the inter-influence between different subsystems in logistics system optimization becomes increasingly significant. As a result, judged from the aspect of logistics system integration, after taking customer's personalized dispatching time into consideration, it's necessary to establish location routing problem with changeable softtime windows model. Based on the feature of the model, this paper adopts hybrid genetic simulated annealing algorithm to gain solution. The experimental result shows that this algorithm is much better than gaining solution solely by adopting hybrid genetic algorithm or simulated annealing algorithm in the aspect of optimal solution, solution quality, calculating efficiency and algorithm stability.

Keywords: hybrid genetic simulated annealing algorithm, logistics system, location routing problem

1 Introduction

The logistics dispatching system mainly includes two parts: the first one is the assignment of dispatching tasks. It's also normally known as the location allocation problem (LAP). The second one is routing between dispatching locations. In other words, it is the vehicle routing problem (VRP). The aforesaid two parts are interdependent and mutual binding. In practice, since both problems need to be taken into comprehensive consideration, the location routing problem is formed.

The location routing problem (LRP) can be described as: after being provided with a serial potential facility points conforming to the actual situation, we need to determine a serial facility locations and a route from different facilities to different customers based on the target of meeting the requirement of the program (shortest road, least expense, least time and least deployed vehicles). Generally speaking, the total cost needs to be minimized.

The location routing problem contains two senses: the location allocation problem is defined as to determine the facility quantity and locations within a geographic range based on geographic distribution and goods allocation relationship. The vehicle routing problem can be defined as the that vehicle designs an optimal goods dispatching route based on one or multiple facility to distributed geographic customer points while meeting a serial of binding conditions within the precondition that the facility location, customer points and route information are known. Based on it, we can design a set of vehicle dispatching rout to meet the target function. Generally speaking, VRP's target function has the least cost.

Under the circumstance of E-commerce, customer has both demand (dispatching) and supply (collecting). As a result, the goods flow is actually bidirectional. In addition, customer has personalized requirement of dispatching time. Based on aforesaid features, the location route problem can be extended to below problems:

Location Routing Problem with Changeable SoftTime Windows (LRPCSTW) can be described as below: to design a serial of facility locations based on the given potential facilities corresponding to actual issues and determine a route from different facilities to different customer points, so that they could try to arrive at all customer points within the given time; otherwise, loss will be contributed by stopping the vehicle for waiting for delay in dispatching. The target is to design the route of minimized cost. LRPCSTW, in addition to fulfilling the classic LRP requirements, also needs to consider visiting time limitation to find out the rational solution.

2 The basic elements of genetic algorithm

The basic operations of GA include encoding, appearance of initial population, fitness calculation, selection, crossover and mutation.

2.1 GENETIC CODE

According to the workflow of GA, when using GA in solving problems, a relationship should be established between the actual presentation of target problems and the bit-string structure of the chromosome in GA, namely the encoding and decoding operations should be determined. The encoding is to express the solutions with a code so as

* *Corresponding author* e-mail: wangchengduan@163.com

to make the problem state space corresponding to the coding space of GA, which relies heavily on the nature of the problems and which will affect the design of genetic manipulation. The optimization of GA is carried out in the code space corresponding to certain encoding mechanism instead of working directly on the parameters of the problems; therefore, the selection of the code is an important element affecting the algorithm performance and efficiency [1]. In function optimization, different code lengths and code systems place a great influence on the accuracy and efficiency of the problems. Binary encoding demonstrates the solutions to the problems with a binary string while decimal encoding presents the solutions with a decimal string. Obviously, the code length will affect the algorithm accuracy and the algorithm will give out larger memory space. Real-number encoding is to show the solutions with a real number and it has been extensively applied in high-dimensional and complex optimization space since it has solved the influence played by encoding on the algorithm accuracy and memory space. For the given optimization problem, the space formed by GA phenotype collection individuals is called problem space while that consisted by GA genotype individuals is called GA coding space. The genetic operators are implemented in the bit-string individuals in GA coding space [2].

2.2 GENETIC OPERATOR

The operators of the standard genetic algorithm often include three basic forms: selection, crossover and mutation, which make up the core that GA has strong search capacity and which are the main carriers of the reproduction, hybridization and mutation produced in the simulation of the natural selection and the genetic process. GA realizes the group evolution by using the genetic operators to reproduce a new generation of groups and the design of the operators is not only a key component of the genetic strategy, but also a basic tool to adjust and control the evolution process [3]. This paper will discuss the effect the genetic operators play on the convergence separately, which helps to learn about the characteristics and importance of genetic operators better.

2.2.1 Selection Operator

Selection is to choose the individuals with high fitness value from the current group to produce the matingpool and it mainly includes fitness-proportionate selection, Boltzmann selection, rank selection, tournament selection and elite-preserving selection. In order to prevent the optimal individuals of the current group from losing in the next generation due to selection errors or the destructive effects of crossover and mutation, DeJong has come up with the elitist selection. In addition, Holland and others have also brought forth steady-state selection. The selections operators are mainly used to prevent gene deletion and improve the global convergence and the calculation efficiency and the most commonly-used

selection operators are fitness-proportionate selection operator and the elite-preserving selection operator.

Proportional model, also called Roulette wheel, is a method of playback random sampling and its basic idea is that the probability of every operator to be selected is directly proportional to its fitness. Because of random computation, the selection error of this method is so big that some individuals with high fitness fail to be selected; however, this is still one of the commonly-used selection operators.

Assume that the group size is M and the fitness of the individual i is F_i. Then the probability p_i of the individual i to be selected is [4]:

$$p_i = \frac{F_i}{\sum\limits_{i=1}^{M} F_i}, (i = 1, 2, \cdots, M).$$

In running GA, new individuals emerge continuously from such genetic operations as crossover and mutation on the individuals. Although more and more excellent individuals will appear in the group evolution, they may destroy the individuals with optimal fitness due to the randomness of selection, crossover and mutation. We hope that the individuals with optimal fitness can be preserved till the next-generation group as much as possible; therefore, we need to apply Elitist Model, meaning that the individuals with the highest fitness in the current group won't participate in the crossover and mutation but replace the individuals with lowest fitness produced by the current group after crossover and mutation.

2.2.2 Crossover Operator

The so-called crossover operation in GA means that two matching chromosome individuals replace part of their genes in accordance with a certain way and form two new individuals. As a significant characteristic of GA, crossover operation plays a key role in GA and it is also a main method to produce new individuals.

Crossover operation is usually divided into the following several steps:
a) Randomly take out a pair of mating individuals from the matingpool;
b) Randomly take one or more integers k from [1, L-1] as the crossover position of the pair of mating individuals according to the bit string length L;
c) Carry out crossover operation according to the crossover probability $p_c (0 < p_c \leq 1)$; the mating individuals replace part of their contents and form a pair of new individuals at the crossover positions [5].

The most commonly-used crossover operator is One-point Crossover, which refers to set a crossover point randomly in the individual encoding string and replace some chromosome in these two mating individuals at this point. One-point Crossover has an important characteristic: if the relationship between the neighbouring loci can provide better individual character and higher

individual fitness, then it will be less possible for this One-point Crossover to destroy such individual character and lower the individual fitness.

It will be faster to solve knapsack problem with and/or swap operation, the specific realization methods of which include:
a) Choose two parent strings A and B according to the roulette wheel selection mechanism;
b) Produce a substring A' from A and B according to logic and operation;
c) Produce a substring B' from A and B according to logic or operation.

2.2.3 Mutation Operator

As a local random search, mutation can avoid the eternal loss of some information caused by selection and crossover operators if combined with these operators. If mutation operation is conducted on the individuals with certain probability instead of single hybridization operation, mutation will randomly change the vectors of the individuals with small probability; in this way, it may result in some new and useful structures may appear and increase the probability to converge to the overall optimization. Mutation operation is a measure to prevent the prematurity of algorithm as well as non-mature convergence. Never take a big mutation rate in the mutation operation. If the mutation rate is bigger than 0.5, GA will degrade into random search and some important mathematic characteristics and search capability will no longer exist [6].

3 The mathematical model of location routing problem (LRP) with changeable SoftTime Windows

3.1 THE CONSTRUCTION OF LRP MODEL WITH CHANGEABLE SOFTTIME WINDOWS

The previous research objective of LRP is to consider cost minimization; however, modern logistics not only consider cost minimization, but also service quality and customer satisfaction of the distribution centre (DC) [7]. Therefore, this paper has also deemed the delivery time restrictions as a part of the optimization objective as well as the cost. LRP with time windows can be divided into hard time window and softtime window. The so-called hard time window refers to a fixed time point while the softtime window is a time period, which can be further divided as fixed softtime window and changeable softtime window. The fixed softtime window means that the start-stop service time intervals required by the customers are the same; on the contrary, the changeable softtime window means that the customers require different start-stop service time intervals. In order to get closer to the practical problems, this paper has investigated the Location Routing Problem (LRP) with Changeable SoftTime Windows (LRPCSTW) and the fundamental objectives of LRP model with changeable softtime window are classified as follows:

1) In order to improve the logistics service quality and customer satisfaction, the cargo is required to be delivered within the required time. Under this circumstance, the top priority of the entire logistics system is how to reasonably arrange for the distribution to meet the time requirements made by the customers;
2) Reasonably plan the driving route with the shortest distribution route and the lowest vehicle-dispatching cost;
3) The entire distribution system realizes total cost minimization to satisfy the high-efficient requirements of the logistics company.

The above objectives form a system and they interact on and affect each other; therefore, LRP is a multi-objective optimization problem.

3.2 LRPCSTW MODEL HYPOTHESIS

Every customer has his/her own requirements on cargo delivery time; there are several distribution centres available; the supply and demand are stable at a certain time; there are sufficient vehicles starting from and stopping at the same centre and the aggregate demand of every service line doesn't exceed the maximum loading of the vehicles and the vehicle models are the same; every customer can only be supplied by one centre; every customer has the same unloading time; select one or several potential distribution centres to complete the specific distribution in every decision and there is only one kind of cargo to be distributed; specially stipulate that the operating range is directly proportional to the transportation cost for the simplified model [8].

3.2.1 The Parameters in the Model

M: There are M customers in this service, who are numbered: 1, 2, , M;
N: There are N potential distribution centres, which are numbered: $M+1,M+2,\ldots,M+N$,
K: There are k callable vehicles, which are numbered: $1,2,\ldots,K$;
K_Q: The maximum load of the kth vehicle;
F_i: The operating cost of the distribution centre i;
C_{ijk}: The unit transportation cost of the vehicle k from the customer i to customer j;
C_k: The unit cargo loading cost of the kth vehicle;
q_i: The cargo quantity demand of customer i;
d_{ij}: The distance from customer i to customer j;
t_{ijk}: The earliest cargo arrival time of vehicle k from customer i to customer j;
LT_i: The earliest cargo arrival time required by customer i;
RT_i: The latest cargo arrival time required by customer i;
T_i: The time when the vehicle arrives at the customer i, $LT_i \leq T_i \leq RT_i$.

3.2.2 The Decision Variables in the Model

$$S_k = \begin{cases} 1, \text{Vehicle} k \text{starts from the distribution center}, k \in K \\ 0, \text{Otherwise} \end{cases}$$

$$X_{ijk} = \begin{cases} 1, \text{Vehicle } k \text{ drives from customer } i \text{ to customer } j, \\ 0, \text{Otherwise} \end{cases} \quad i, j \in M, k \in K, i \neq j$$

$$Z_i = \begin{cases} 1, \text{Distribution center } i \text{ opens}, i \in N \\ 0, \text{Otherwise} \end{cases}$$

3.3 THE CONSTRUCTION OF LRPCSTW MATHEMATICAL MODEL

Objective Function:

$$f_1 = \min \sum_{i=1}^{M} \left\{ \max\left[(LT_i - T_i), 0\right] + \max\left[(T_i - RT_i)0\right] \right\}, \quad (1)$$

$$f_2 = \min \left[\sum_{i=1}^{N} F_i Z_i + \sum_{k=1}^{K} \sum_{i=1}^{M+N} \sum_{j=1}^{M+N} C_{ijk} q_i d_{ij} + \right.$$
$$\left. + \sum_{k=1}^{k} C_k \sum_{i=1}^{M+N} \sum_{j=1}^{N} X_{ijk} q_i \right]. \quad (2)$$

Constraint Conditions:

$$\sum_{i=1}^{M} \sum_{k=1}^{K} X_{ijk} q_i \leq Q_k, \forall j \in M, \quad (3)$$

$$\sum_{i=N+1}^{M} \sum_{k=1}^{M} X_{ijk} - Z_i \geq 0, \forall k \in K, \quad (4)$$

$$\sum_{i=N+1}^{K} \sum_{J=1}^{M} X_{ijk} \leq 1, \forall k \in K, \quad (5)$$

$$\sum_{k=1}^{k} \sum_{i=1}^{M} X_{ijk} = 1, \forall j \in M, \quad (6)$$

$$\sum_{i=1}^{M+N} X_{ihk} - \sum_{i=1}^{M+N} X_{hjk} = 0, \forall k \in K, \quad (7)$$

$$\sum_{i=M+1}^{M+n} \sum_{n=1}^{N} X_{ink} + \sum_{n=1}^{N} \sum_{j=M+1}^{M+N} X_{njk} \leq 1, \forall k \in K, \quad (8)$$

$$\sum_{k=1}^{K} X_{ijk} + Z_i + Z_j \leq 2, \forall i, j \in N, \quad (9)$$

$$X_{ijk}(T_i - T_j) \leq 0, \forall k \in K; i, j \in M, \quad (10)$$

$$T_j = \sum_{i=1}^{M} \left[S_{ijk} \sum_{i=M+1}^{M+N} \sum_{o=1}^{M} \cdots \sum_{i=1}^{M} \left(S_{ink} t_{ink} + S_{olk} t_{olk} + \ldots + S_{ljk} t_{ljk} \right) \right]. \quad (11)$$

Equation (1) means that the cargo must be delivered within the required time by the customer; Equation (2) shows that the total cost minimization to complete this distribution is made up of three parts: the operating cost of the opened distribution centre; the transportation cost of

the vehicles and the unit loading cost of the vehicles in the distribution centres; Equation (3) demonstrates that the aggregate demands of the same vehicle won't exceed the maximum load of this vehicle; Equation (4) makes sure that every opened distribution centre will have callable vehicles; Equation (5) means that the same vehicle can only be called by one distribution centre; Equation (6) means that one customer can only be served by one vehicle; Equation (7) stands for the connectivity of the routes between the customers; Equation (8) means that any two distribution centres won't be at the same route; Equation (9) means that any two distribution centres won't connect each other; Equation (10) refers to the route succession of the vehicles; Equation (11) gives the Equation of T_j in the objective Equation (1) [9, 10].

4 Solve LRPCSTW problems through hybrid genetic stimulated annealing algorithm

To maintain the population diversity is a solution to overcome the premature convergence of the genetic algorithm, which makes the genetic algorithm explore the search region in succession in the evolution. The thought of niche technology comes from the fact that the creatures usually live and copulate with their own species for reproduction. Niche technology is an organizing function to gather the groups with similar biological features and characters together in certain environment and to separate those groups with different features and characters. The organization with the same features and characters is called a species [11]. Niche technology is similar to preserve the excellent individuals in different species and conduct such selection operation, crossover operation and mutation operation as choice mechanism and crowing mechanism in the population or between the populations. The improved genetic algorithm with niche technology can effectively protect the population diversity.

4.1 THE DETERMINATION OF FITNESS FUNCTION

Fitness function plays a significant role in the entire division, which directly affects the quality of the optimal solution. Used to evaluate the individual adaptive capacity, fitness function increases the probability that the individuals with strong adaptive ability to reproduce a next generation and the weak individuals to reproduce another generation. Based on this idea, this paper adopts penalty

function method in constructing fitness function and the purpose to use penalty on the invalid individuals is to reduce its probability to be reproduced to the next generation in calculating the individual fitness. The invalid individuals here refer to those individuals, which fail to meet the constraint conditions. The constraint condition of this paper is the optimal price at time limit and the invalid individuals exceed the maximum constraint. After analysing the thought of penalty function, this paper constructs the penalty item by adopting $Time(\text{S}) < T_{\max}(S)$. In addition to considering the inconsistency in magnitude between the consumption time and price in constructing objective function and the unifying the magnitude, this paper uses normalization factor just as Equation (12) and Equation (13).

$$\sigma_p = PH - PS , \tag{12}$$

$$\sigma_t = \max\left(TS - T_{\max}, T_{\max} - TH\right) . \tag{13}$$

The objective function in this paper is as the following formula:

$$OBJECT = \alpha \cdot e^{\frac{Time - T_{\max}}{T_i \sigma_t}} \cdot \frac{\left|Time - T_{\max}\right|}{\sigma_t} + \beta \cdot \frac{price}{\sigma_p} . \tag{14}$$

This paper discusses minimization problem; therefore, the optimum of this algorithm means the maximum fitness function corresponds to the minimum value of the objective function. The fitness function in this paper is seen as the following Equation:

$$Fitness = \frac{1}{1 + OBJECT} . \tag{15}$$

Time (S) in Equation (14) stands for the aggregate consumption time of the system; Price (S) refers to the aggregate consumption cost of the system and $T_{\max}(\text{x})$ means the time limits of the system. In this formula, $e^{Time - T_{\max}/T_i\partial_t}$ is the penalty item of the invalid individuals. Obviously, only when $Time(\text{S}) - T_{\max} > 0$, $e^{Time - T_{\max}/T_i\partial_t} > 1$; then the objective function in Equation (14) increase while its corresponding fitness function in Equation (15) decreases; therefore, it effectively reduces the fitness value of the invalid individuals. When $Time(\text{S}) - T_{\max} < 0$, $e^{Time - T_{\max}/T_i\partial_t} < 1$, then according to Equation (14), the objective function reduces while the corresponding fitness function increases in the corresponding Equation (15); thus greatly improving the fitness value of the effective individuals. T_i in Equation (14) is similar to the annealing system in SA. Initialize $T_0 = 1$ and $T_{i+1} = 0.98T_i$. This paper makes $\alpha = 0.6, \beta = 0.4$ when $\alpha > \beta$.

4.2 GENETIC MANIPULATION

The selection of genetic operation attaches significant importance to GA efficiency and solution quality and they

are also the core of GA. The genetic selecting operation, crossover operation and mutation operation simulate the multiply, hybridization and mutation in the nature.

Adopt Boltzmann selecting operation on the problem investigated in this paper and the probability to select the individual is as follows [12, 13].

$$P_s\left(S_i\right)\frac{1}{\sum_{i=1}^{n} e^{f(S_i)}/T} , i = 1, 2, ..., n , \tag{16}$$

where $T > 0$ and T refers to the temperature parameter in the annealing process.

Crossover operation uses ingle-point crossover and reverse operation is also adopted in this process. In this way, it can make the important genes more compact, which is equal to redefine the gene block in the algorithm. See the example of this operation is as follows:

As for the binary string with a length of 10, mark the important genes with underline:

$\underline{1}\ \underline{0}\ \wedge\ 1\ 1\ 1\ 0\ 1\ \underline{1}\ \wedge\ 0\ 1$

$\left(\text{And}\ \wedge\ \text{is the inverse transposition}\right).$

Produce new binary string after reverse operation:

$\underline{1}\ \underline{0}\ \underline{1}\ 1\ 0\ 1\ 1\ 1\ 0\ 1 .$

From the new binary string, it can be seen that the important genes are more compact, thus reducing the probability that the important genes will be dispersed in the single-point crossover.

Mutation operation: In the genetic algorithm, mutation operation is usually realized by mutating the binary string according to the mutation probability P_m and SA annealing temperature is used to decide the selection of mutation probability with the mutation probability: P_m=0.01.

4.3 THE IMPLEMENTATION PROCESS OF THE ALGORITHM

Step1: Initialize the population and finish the coding of the solution space, $P(0) = \{S_1, S_2 \cdots, S_z\}$.

Step2: Calculate the fitness value of every individual in the initial population and according to the fitness calculation formula, namely Formula (15), determine the initial temperature of the simulated annealing algorithm, namely $T(0) = (f_{\min} - f_{\max})/\text{In}p_r$. In the initial temperature formula, f_{max} is the largest fitness value in the initial population; f_{min} is the smallest fitness value in the initial population; P_r is the worst initial acceptance probability and in this paper, P_r=0.7.

Step3: Conduct genetic operation to the population $P(t)$: select single-point crossover operation and mutation operation by adopting Boltzmann in this paper and produce the new population $P(t+1)$. The possibilities to make single-point crossover and mutation operation are P_c=0.8 and P_m=0.01 respectively.

Step4: Produce a new-generation population $P(t+2)$: According to Metropolis criterion, randomly disturb the neighbourhood region of the individual S_i in the population of $P(t+1)$. And $\Delta f = fit(S_i) - fit(S_j)$, Compare Δf with 0. When $\Delta f \leq 0$, directly copy S_j to $P(t+2)$; if not, produce a random number r among [0, 1], directly copy S_j to $P(t+2)$ when $r < e^{(-\Delta f/T(N))}$; in other cases other than the above two circumstances, copy S_i to $P(t+2)$.

Step5: Conduct elimination: calculate the Euclidean Distance between individual S_i and individual S_j in $P(t+2)$,

namely $\left\| S_i - S_j \right\| = \sqrt{\sum_{k-1}^{M}\left(S_{ik} - S_{jk}\right)^2}$; among them,

$i = 1, 2, ..., n-1, j = i+1, ..., n$ and M means the quantity of

the decision variables in the problems to be settled. In the formula $\left\| S_i - S_j \right\| \leq L$, when L stands for distance parameter, calculate the fitness value of individual S_i and individual S_j and abandon the individuals with smaller fitness value and do not operate when $\left\| S_i - S_j \right\| > L$.

Step6: When Q arrives at the maximum value, the program stops running and quits; otherwise, use the corresponding cooling means to lower the annealing temperature. This paper adopts rapid annealing method with the Equation $T(N) = T_0 / 1 + \alpha N$, where $\alpha = 0.9$. Make $N=N+1$, $t=t+2$ and repeat **Step 3**. The flow chart of genetic simulated annealing algorithm is as follows [14]:
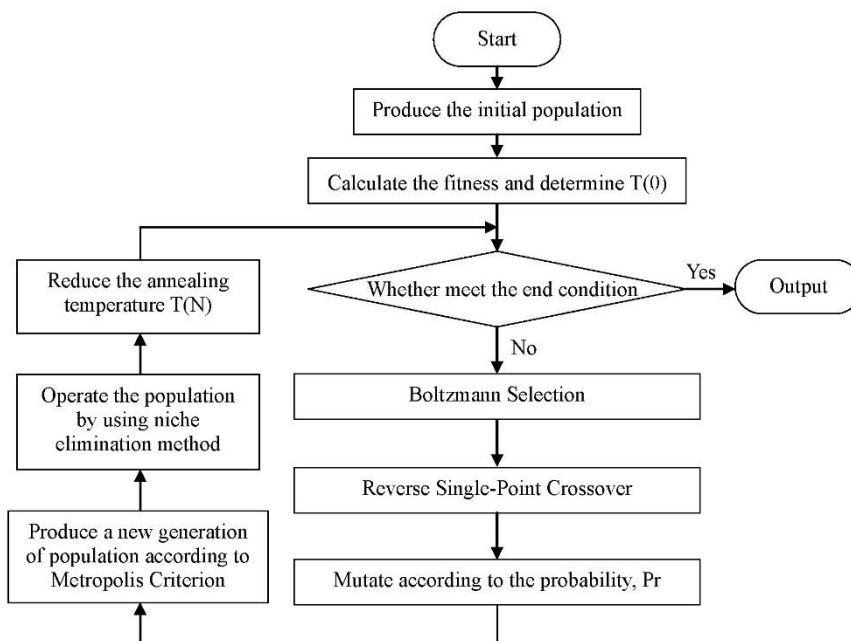
FIGURE 1 Flow Chart of Genetic Simulated Annealing Algorithm

5 Gaining solution with hybrid genetic simulated annealing algorithm

5.1 AN EXPERIMENTAL ANALYSIS ON THE LRP WITH CHANGEABLE SOFTTIME WINDOWS

There are 2 third-party logistics enterprises locating at (38km, 70km) and (55km, 80km) respectively. Both enterprises are able to provide 54t dispatching capacity independently and fulfil 30 customers' requirements solely. There are 5 potential dispatching canters and 30 customers with time requirements.

The five potential dispatching canters can offer some vehicles with 2 kinds of dispatching capacities, namely vehicle A of 10t dispatching capacity and vehicle B of 8t dispatching capacity. The fixed vehicle cost g=10 RMB. The unit traveling cost of 8t vehicle is C_{ij}=1.5 RMB/km, while the unit traveling cost of 10t vehicle is C_{ij}=1.6 RMB/km. The maximum traveling capacity of both kinds

of vehicles is 100km. The unit delay expense CCS_k^l of 8t vehicle is 50 RMB/h, while the unit delay expense CCS_k^l of 10t vehicle is 60 RMB/h. the waiting cost CS_k^l of 8t vehicle is 40 RMB/h, while the waiting cost CS_k^l of 10t vehicle is 50 RMB/h. The five potential dispatching centre has the same capacity, T_r=35t, while the cost of establishing and running a potential dispatching centre is F_r=300. The unit dispatching cost between enterprises and dispatching centre is c_r=0.15RMB/$t.km$. The average speed of vehicle is 20km/h.

How to make rational arrangement of enterprise quantity and location, dispatching centre quantity and location and vehicle route to minimize the total logistics chain while fulfilling customer's time requirements.

5.2 GAINING SOLUTION WITH HYBRID GENETIC SIMULATED ANNEALING ALGORITHM

Adopt below parameters: group scale N=80, maximum iteration time max_{gen}=300, cross operator p_c=0.90, mutation operator p_{m1}=0.2 and p_{m2}=0.01, initial temperature T_0=250 and temperature dropping coefficient δ=0.90. Adopt C programming Language to fulfil the above mentioned algorithm on a computer with CPU 1.8GHz and 512M memory and solve the problem randomly for 10 times.

Judged from the result, the solutions gained from the hybrid genetic simulated annealing algorithm for 10 times are all of high quality solutions. The average total traveling distance is 480 km and average total cost is 1926.6 RMB. The calculation results are stable while the total distance of worst solution is only 6.6% longer than that of best solution, and the total cost of the worst solution is only 2.5% more than the best solution. In the aspect of

calculation efficiency, the average iteration times of 10 calculations is 247.1. As a result, the efficiency is significantly high.

Among the solutions, the 5th calculation gains the solution among the 10 calculations: Enterprise 1 locating at (38km, 70km), by deploying dispatching centre 2, 4 and 5 to arrange dispatching route with time window for 30 customers, only needs to traveling 470km with a total cost of 1905.5 RMB.

Its searching process is indicated as Figure 2. Judged from the figure, this algorithm not only has fast convergence speed but also has outstanding global searching ability. The global most optimal gene protection strategy, use of self-adaptation mutation operation, Boltzmann mechanism from simulated annealing algorithm, control of cross and mutation operation in genetic algorithm are helpful in breaking the local limit for the sake of finding more optimal solutions.
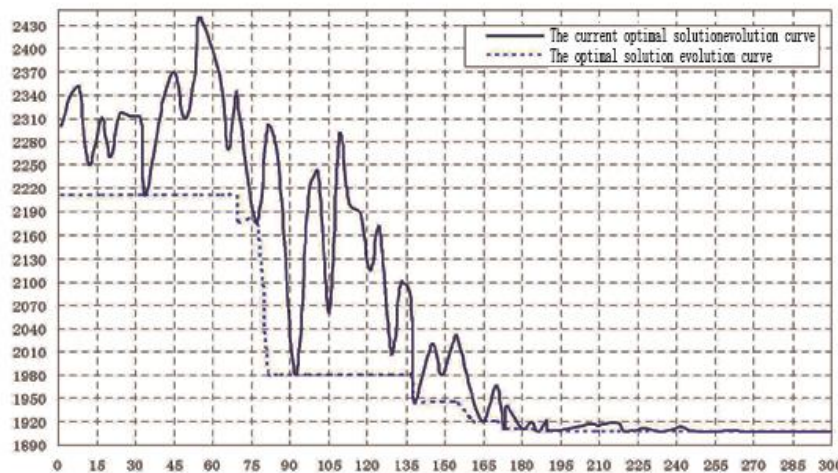
FIGURE 2 Evolutionary curve of LRPCSTW by hybrid genetic simulated annealing algorithm

TABLE 1 Final optimization of LRPCSTW by hybrid genetic simulated annealing algorithm

Dispatching centre	Location route arrangement schedule(h)	Dispatching amount(t)	Vehicle Model/(t)	Traveling Capacity(km)	Total distance(km)
2	0-11-10-9-8-7-6-5-0 0-0.55-0.8-1.85-2.1-3.0-3.35-3.65-4.8	9.0	A /10.0	96	195
	0-13-12-30-0 0-0.75-1.05-1.5-2.75	7.0	B /8.0	55	
	0-17-14-15-16-0 0-0.5-0.75-1.25-1.4-2.2	8.0	B /8.0	44	
5	0-29-24-23-0 0-0.25-0.65-1.1-1.85	6. 0	B /8.0	37	137
	0-22-21-20-19-18-0 0-1.6-2.1-2.4-2.8-3.05-5.0	8.0	B /8.0	100	
4	0-1-3-4-2-0 0-0.25-1.5-2.2-3.65-3.8	8.0	B /8.0	76	138
	0-27-25-26-28-0 0-1.25-1.6-2.05-2.3-3.1	8.0	B /8.0	62	
Total distance (km)		470			

TABLE 2 Final total cost of LRPCSTW by hybrid genetic simulated annealing algorithm

Dispatch centre	Fixed cost of vehicle	Fixed cost of dispatching centre	Delivery Cost between dispatching centre and customer	Delivery Cost between enterprise and dispatching centre	Total cost
2	30	300	302.1	144.4	776.5
5	20	300	205.5	55.9	581.4
4	20	300	207	20.6	547.6
total cost			1905.5 (RMB)		

5.3 THE COMPARISON OF THREE ALGORITHMS

The hybrid genetic simulated annealing algorithm is better than sole use of hybrid genetic algorithm or simulated annealing algorithm in the aspect of optimal solution, solution quality, calculation efficiency and algorithm stability.

TABLE 3 Comparison among hybrid genetic simulated annealing algorithm, GA and SA

Algorithm	Hybrid genetic simulated annealing algorithm	Genetic algorithm	Simulated annealing algorithm
Total distance(km)	480	517.5	498.5
Standard deviation of solution	16.12	18.45	20.03
Average total cost (RMB)	1926.6	1981.3	1997.2
Standard deviation of solution	21.75	81.46	105.57
Average vehicle quantity	7.2	7.6	7.8
The first time of gaining the most optimal solution	186	237	329
The most optimal total distance (km)	470	498	479
The most optimal total cost (RMB)	1905.5	1953.8	1963.2

6 Conclusions

This paper, from the aspect of logistic system integration, establishes a model for location route problem with changeable softtime windows. These models are effective in indicating the diversified feature of dispatching service under the circumstance of E-commerce as well as fulfilling customer's personalized and diversified requirement. For the location route problem with changeable softtime windows, this paper designs the hybrid genetic simulated annealing algorithm in gaining solution. The simulated calculation result shows that the hybrid genetic simulated annealing algorithm designed by this paper is better than sole use of hybrid genetic algorithm or simulated annealing algorithm in the aspect of optimal solution, solution quality, calculation efficiency and algorithm stability.

References

[1] Moon I, Lee J, Seong J 2012 Vehicle routing problem with time windows considering overtime and outsourcing vehicles *Expert Systems with Applications* **39**(18) 13202-13

[2] Chiang T 2013 Enhancing rule-based scheduling in wafer fabrication facilities by evolutionary algorithms: Review and opportunity, *Computers & Industrial Engineering* **64**(1) 524-35

[3] Gelareh S, Merzouki R, McGinley K, Murray R 2013 Scheduling of Intelligent and Autonomous Vehicles under pairing/unpairing collaboration strategy in container terminals *Transportation Research Part C: Emerging Technologies* **33** 1-21

[4] Yu J, Dong Y 2013 Maximizing profit for vehicle routing under time and weight constraints *International Journal of Production Economics* **145**(2) 573-83

[5] Marinakis Y, Marinaki M 2014 A Bumble Bees Mating Optimization algorithm for the Open Vehicle Routing Problem *Swarm and Evolutionary Computation* **15** 80-94

[6] Corman F, D'Ariano A, Pacciarelli D, Pranzo M 2014 Dispatching and coordination in multi-area railway traffic management *Computers & Operations Research* **44** 146-60

[7] Muñuzuri J, Grosso R, Cortés P, Guadix J 2013 Estimating the extra costs imposed on delivery vehicles using access time windows in a city *Computers, Environment and Urban Systems* **41** 262-75

[8] Vidal T, Crainic T D, Gendreau M, Prins C 2013 Heuristics for multi-attribute vehicle routing problems: A survey and synthesis *European Journal of Operational Research* **231**(1) 1-21

[9] Tu W, Fang Z, Li Q, Shaw S, Chen B 2014 A bi-level Voronoi diagram-based metaheuristic for a large-scale multi-depot vehicle routing problem *Transportation Research Part E: Logistics and Transportation Review* **61** 84-97

[10] Lin C, Choy K L, Ho G T S, Chung S H, Lam H Y 2014 Survey of Green Vehicle Routing Problem: Past and future trends *Expert Systems with Applications* **41**(4) Part 1 1118-38

[11] Ferrucci F, Bock S 2014 Real-time control of express pickup and delivery processes in a dynamic environment *Transportation Research Part B: Methodological* **63** 1-14

[12] Wu Y, Luo J, Zhang D, Dong M 2013 An integrated programming model for storage management and vehicle scheduling at container terminals *Research in Transportation Economics* **42**(1) 13-27

[13] Pillac V, Guéret C, Medaglia A L 2012 An event-driven optimization framework for dynamic vehicle routing *Decision Support Systems* **54**(1) 414-23

[14] Lacomme P, Larabi M, Tchernev N 2013 Job-shop based framework for simultaneous scheduling of machines and automated guided vehicles *International Journal of Production Economics* **143**(1) 24-34

Author

Chengduan Wang, born in March, 1967, Weifang, China

Current position, grades: Professor, Dean at the school of computer engineering.
University studies: M. Sc. degree in computer applications from Shandong Science and Technology University (2007).
Scientific interest: intelligent computing, software engineering.
Publications: 5 papers, over 5 books.

Operation Research and Decision Making

Research on computer network English language learning

Hongxia Wei[*]

Foreign Languages School of Anhui Polytechnic University, Wuhu, Anhui 241000, China

Abstract

Network information technology affects English language learning mode all the time with the rapid development of information technology. University English course teaching also turns from traditional classroom to modern network teaching. As a bran-new teaching form, university English teaching brings not only advantage for classroom teaching but also impact on teaching idea, teaching means, teaching pattern, etc. It breaks through the ecological balance of traditional university English course and affects the successful implementation of teaching reform mode. Aiming at improving university English course teaching quality and perfect teaching reform mode, this paper studied new English teaching means based on computer network technology, analysed imbalance of university course teaching in teaching reform mode and then strove to construct a frame of ecologicalization university English course teaching.

Keywords: network environment, university English teaching, voice recognition confidence level, digital game teaching

1 Introduction

Rapid development of modern information technology brings huge change in the world such as life, work and communication mode. It affects education quality with its flexibility, convenience and effectiveness. Internet provides unprecedented learning opportunity for learners. People can not only share study resource but also overcome the obstacle of space and time, which brings new prospective for education. Education idea changes a lot with the rapid development of network information technology. Nowadays, rapid development of education in our country aims at promoting education modernization by network information technology [1]. National Medium and Long-term Plan for Education Reform and Development (2010-2012) clearly points out that the revolutionary effect of information technology on education development must be paid high attention to and education reform should go on information development way from the macroscopic level. Therefore, many domestic scholars devote themselves to the research on this aspect. Luo Hongwei from East China Normal University constructed network English teaching emotional design frame and conducted empirical research so that solved athymia of network English teaching based on soft systems methodology and instructional design theory in Research on Emotional Design of Network English Teaching [2]. Chen Jing from Suzhou University analysed relative concept and theoretical basis of autonomous learning and ability composition of autonomous learning in network environment combining with the characteristics of network environment in Research on English Autonomous Learning of Secondary Vocational Students under Network Environment [3]. That paper systematize scattered knowledge and form relative

complete knowledge content system, which provides guidance for network English autonomous learning of secondary vocational college students. Ma Chongyu and Yang Gang made a comprehensive remark on historical development and foreign and domestic research overview of 3-D virtual language learning environment and discussed enlightenment and development prospect on foreign language education in New Direction of Computer Assisted Language Learning Development - 3-D Virtual Language Learning Environment [4]. It is not hard to find that these researches are prompting English teaching in our country to move towards informatization.

This paper is about optimal frame of university English class teaching in network environment supplemented by bran-new teaching mode based on computer network technology under the background of university English teaching reform pushed by Education Ministry. It emphasized construction of autonomous learning strategy and teacher development idea in the environment of network. Its achievement has certain practical value and practical significance for construction of ecological system balance of informationization foreign language teaching and improvement of university English teaching reform effect.

2 Main thought of research

Overall, foreign language system is a natural and open large ecological system composed by factors and environment within a certain space according to education ecology theory. In university English teaching ecological system, teaching factors affect and depend on each other in existing process. Although integration of information technology and university English teaching has advantage, imbalance of university English teaching ecological

---

[*]*Corresponding author* e-mail: whxahgf@163.com

system appears in integration process. The key point of construction of dynamic balanced university English class teaching ecological system is to discuss how factors in that system own specific reasonable dynamic ecological niche and achieve harmony [5]. This paper would construct a university English teaching ecological system according to

optimality principle to improve university English reform effect in the perspective of ecology in the aspects of student, teacher and teaching environment. The following Figure 1 shows optimal frame conception of university English course teaching under network environment.
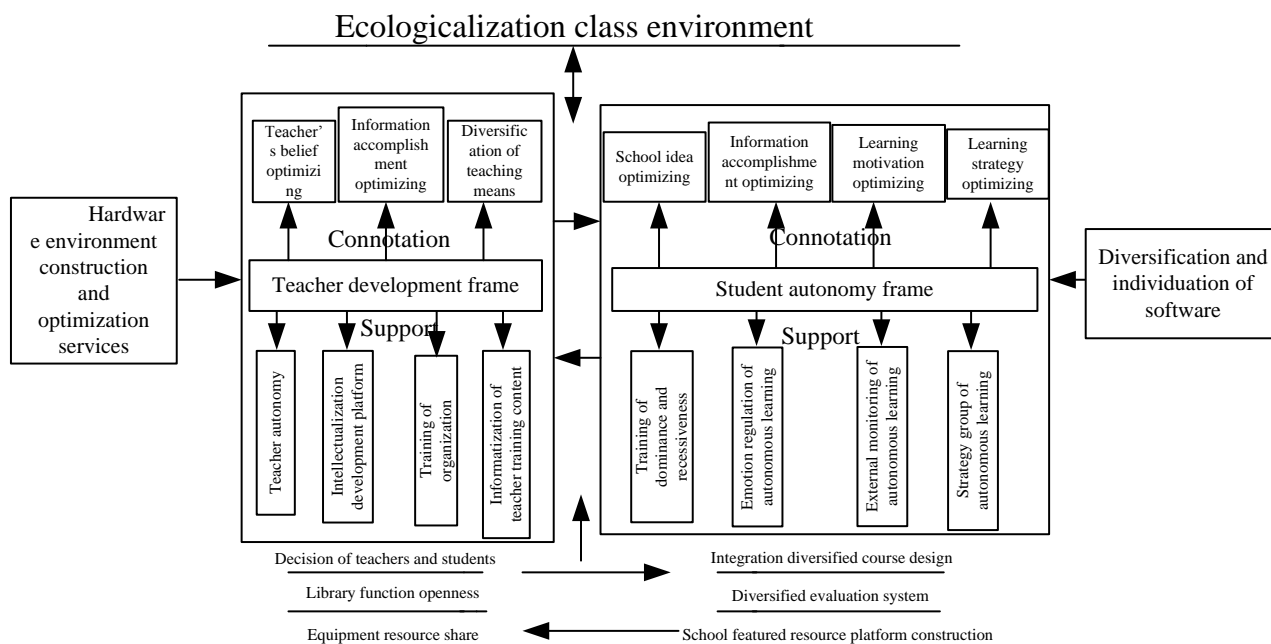
FIGURE 1 Optimal frame of university English class teaching under network environment

3 Main content of research

3.1 INTEGRATION NATURE OF INFORMATION TECHNOLOGY AND UNIVERSITY ENGLISH COURSE

Nature and connotation of information technology and university English course integration is under the guidance of advanced foreign language teaching theory. In the process of English teaching, we should apply computer network technology scientifically, promote students' cognition, motivate students' emotion, enrich teaching environment and integrate various teaching resource and teaching factors to make university English teaching system generate aggregation effect in technology application, improve university English education quality and realize ultimate goal of teaching reform [6]. Integration means that information technology will become an organic component of teaching. It is not only the tool of teaching but also the driving force of career development of university English teacher, autonomous learning of students and cultivation of creative thinking

ability. And technology also becomes important component of teaching evaluation and management.

3.2 UNIVERSITY ENGLISH CLASS TEACHING OPTIMIZATION THEORY AND PRINCIPLE UNDER NETWORK ENVIRONMENT

University English class teaching optimization theory is an education ecological theory based on ecology. Its nature is to study relationship between student, teacher and factors as well as their relationship with environment applying theory and principle of ecology, think and explain imbalance phenomenon and reason in class teaching in the perspective of education ecology, construct university English class teaching ecological system under network environment by ecological principles and optimize university English course teaching. Optimality principle of university English course teaching is explored out combining with university empirical research under the guidance of education ecology theory. Its main content is based on the following five principles, as shown in Table 1.

TABLE 1 University English class optimality principle

Principle	Main content
Optimality principle	We should treat the overall process of teaching by systematic means. All components and teaching factors in and outside class contained in teaching system are all connected. We should choose optimal teaching means, technique, content and form and form optimal teaching construction.
Principle of "stabilizing teaching structure, compatible with teaching factor"	Education ecological theory emphasizes the relationship between stability of teaching factors and balance of teaching environment and the relationship between compatibility of class teaching factors and ecological environment harmony. The objective of stability need to be realized by compatibility technique and means
Principle of "restricting teaching operation, promoting individual development"	"Restriction" is the technique to maintain ecological system stability. "Promotion" is the objective of university English class teaching optimization. In the perspective of education ecology, university English class teaching factors have their specific function. All factors operate on their own space and time position in the meantime and play unique role in their ecological niche.
Principle of "autonomous learning of dominate type"	Students are encouraged to construct their own knowledge frame and turn from passive knowledge receivers to positive knowledge constructors.
Principle of "diversified interactive teaching"	Refer to multidirectional and multilevel interactive teaching mechanism between teacher and student, student and student, human and machine under network environment.

3.3 SCHEME OF UNIVERSITY ENGLISH CLASS TEACHING OPTIMAL FRAME

We constructed a university English class teaching optimal frame centred on teacher development and student autonomy and supported by ecological class teaching environment under the guidance of education ecology theory and five teaching optimality principles. This frame involves teacher development frame, student autonomous frame and class teaching environment:

1) Teacher development frame refers to constructing modernized teacher development frame centred on optimizing teacher teaching views, teacher information accomplishment and construction of diversified teaching means conception. The specific content of frame is as shown in the following Figure 2:
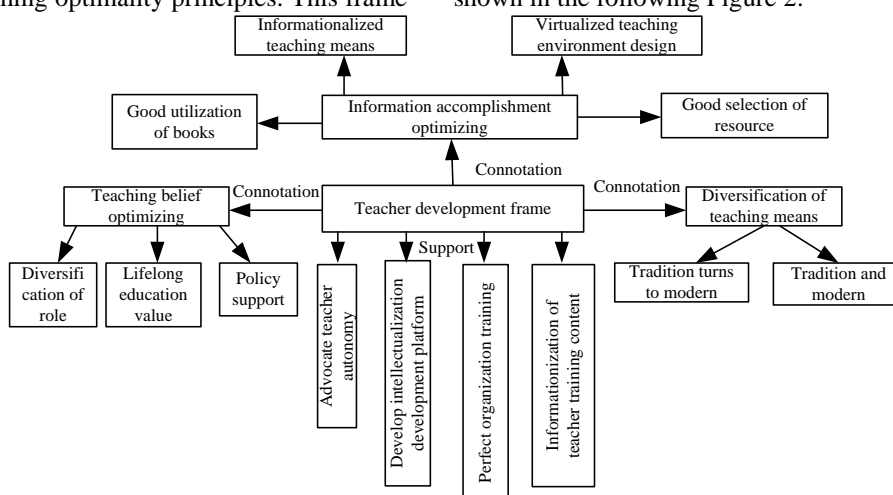
FIGURE 2 Modernized university English teacher development frame

2) Student autonomy frame refers to establish student autonomous learning ability cultivation frame involving optimizing of student autonomous learning value, informational accomplishment, learning motivation and autonomous learning strategy, as shown in the Figure 3:
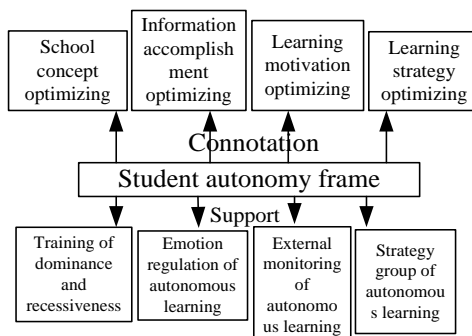
FIGURE 3 Student autonomous learning ability cultivation frame

3) Ecologicalization university English class teaching environment construction refers to university English class environment optimizing frame for optimizing hardware construction and service and constructing diversified and individualized software environment, as shown in the Figure 4:
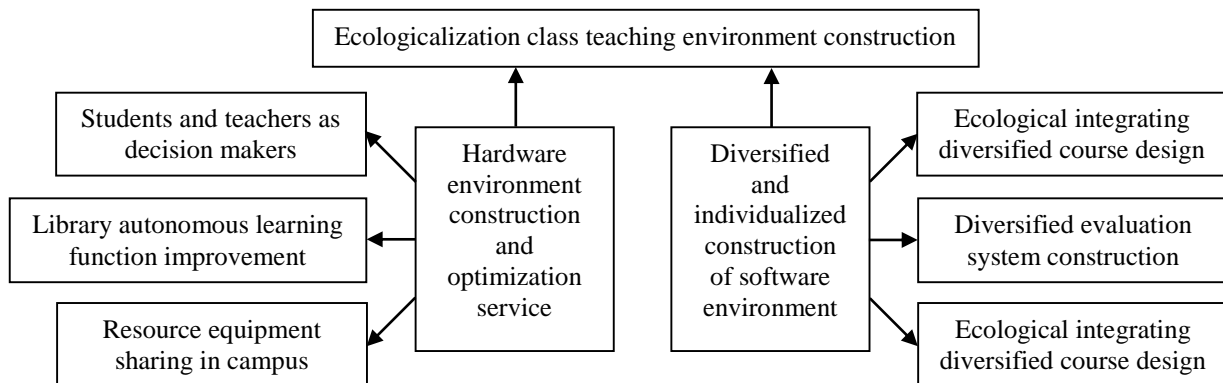
FIGURE 4 Ecologicalization university English class teaching environment

4 Teaching means integrating with computer technology

4.1 INTEGRATION OF DIGITAL GAME AND ENGLISH COURSE

Digital course is a kind of entertainment activity conducting by one or more people with electronic equipment according to certain rule. Nowadays, this kind of game has blended in teaching and an innovated teaching pattern has formed [7]. This pattern is termed as digital game teaching. It regards game as media to transmit teaching content. The application of digital game in English teaching in nature is the specific integration of computer information technology and course in practice. In the integration process, computer steps from secondary to domination. Teachers retreat from teaching centre to assistance while students become the subject and centre of teaching. And the whole teaching becomes a student-oriented, teacher-assisted and computer dominated teaching pattern. Therefore, teachers can make out teaching game with strong pertinence to meet teaching requirement by authoring software and hardware based on teaching content.

4.2 ENGLISH LANGUAGE LEARNING SYSTEM BASED ON CONFIDENCE LEVEL

That system is composed of two independent subsystems: speech recognition and confidence level evaluation, as shown in the following Figure 5. In first stage (speech recognition), we first apply train out all HMM models by voice library and then search out likelihood and residing information by Veterbibeam. In second stage (confidence level evaluation), we should calculate out likelihood ratio from decoding information of the first stage, then make lots of statistics on likelihood combined with pronunciation subjective testing of trainer to confirm an objective distortion degree standard and finally evaluate the correctness of learners' pronunciation on that basis.
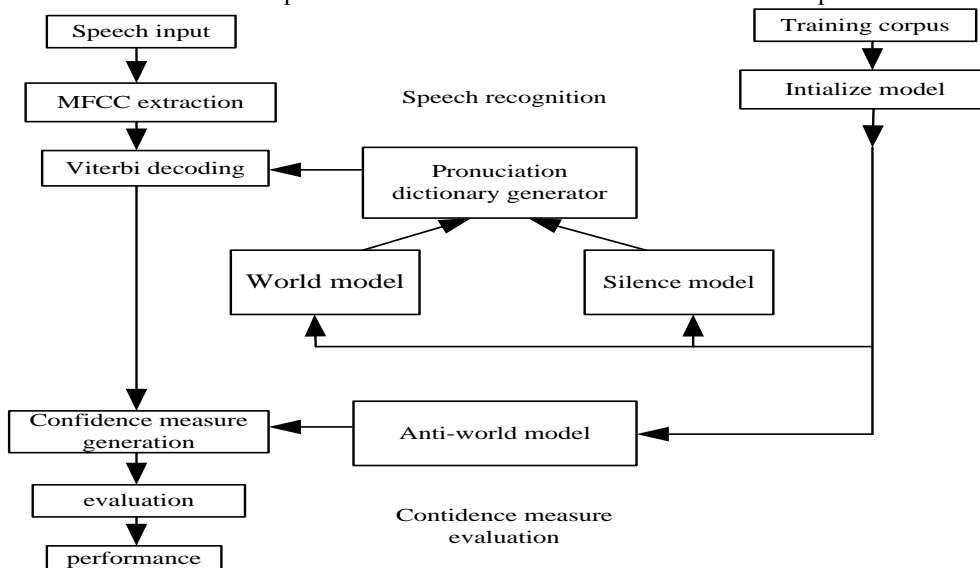
FIGURE 5 English language learning system based on confidence level

Network topology in English language learning system is shown in Figure 6. If we adopt network topology that only contains vocabulary words model, the recognition result will be "sil + in-vocabulary + sil". Therefore, no matter the input voice is vocabulary word, we can always obtain decoding information related to vocabulary words model. We can obtain decoding information related to anti-phoneme words model and anti-vocabulary words model by that method, then transmit this information to confidence level module, and finally evaluate the input voice.
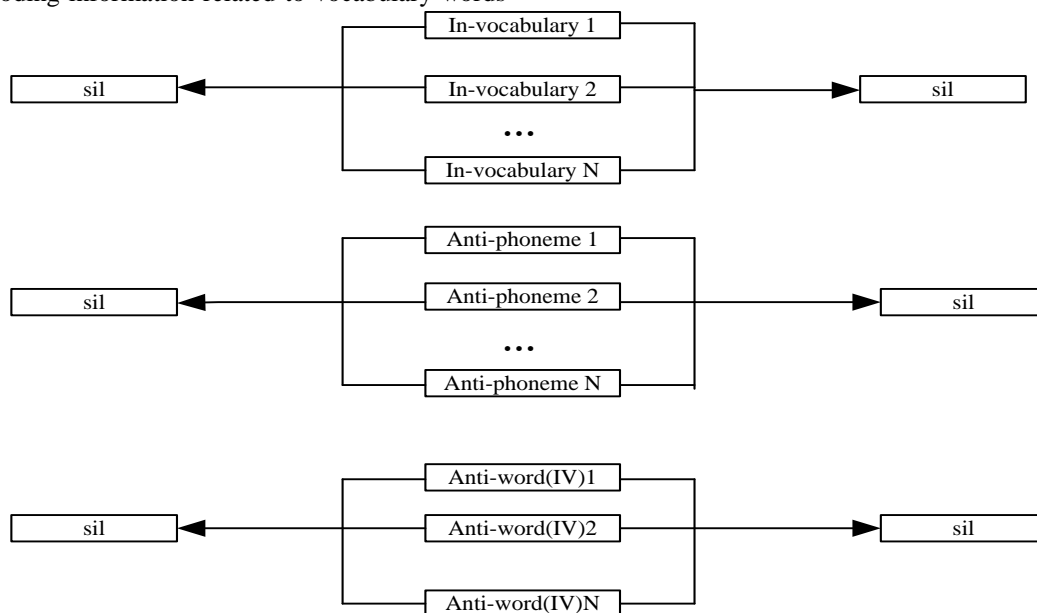
FIGURE 6 Network topology of English language learning system

5 Conclusions

Lots of scholars turn research direction to network foreign language teaching with its development. However, most of them only concern about the relationship between technology and learning. So far, most researches related to integration of network technology and course focus on comparative research on new and old teaching mode and research on construction of informatization teaching mode [8]. This paper is an exploratory research based on the relative researches of Chinese and overseas scholars combined with university English teaching and teaching reform reality based on network environment. It tried to construct an optimal plan for English teaching with integration of traditional teaching course and computer technology in the perspective of theory of ecological education and then studied a new approach for informatization English language learning combined with new computer teaching means. It not only improves imbalance between students, teachers and environment of English course in network environment, makes full use of computer technology for teaching means innovation and improves teaching efficiency, which provides new pathway for later English language learning research.

References

[1] Cai J G 2012 Restudy on university English teaching objective positioning in our country under globalization background *Foreign Language & Foreign Language Teaching* (3) 5-8
[2] Luo H W 2009 Research on emotional design of network English teaching *Shanghai East China Normal University*
[3] Chen J 2011 Research on English autonomous learning of secondary vocational students under network environment *Suzhou University*
[4] Ma C Y, Yang G 2014 New direction of computer assisted language learning development - 3-D Virtual language learning environment *Journal of Hebei University* **39** (1) 152-155
[5] Michael B, Angelene M, Lin Z 2014 It's not that tough: Students speak about their online learning experiences *Turkish Online Journal of Distance Education-TOJDE* **13**(2) 226-241
[6] Yang S H 2009 Using blogs to enhance critical reflection and community of practice *Education Technology & Society* **12** (2) 11-21
[7] Zhao R 2009 Digital game teaching in the perspective of the problems faced by computer network English teaching *Media in Foreign Language Instruction* (125) 69-74
[8] Cai S, Yu S Q 2010 Development trend of 3-D Virtual learning environment in the perspective of sloodle O*pen Education Research* **16**(2) 98-104

Author

Hongxia Wei, born in 1971, Anhui Province of China

Current position, grades: associate professor.
University studies: Bachelor of Arts degree in English Education, Anhui Normal University in 1995, Master of Arts degree of foreign linguistics and applied linguistics, Nanjing Normal University in 2010.
Scientific interest: English linguistics.

A research about the predictive control of dynamic feedforward neural network based on particle swarm optimization

Lizheng Liu[1, 2*], Fangai Liu[2], Feng Yang[1]

[1]*Shandong University of Finance and Economics, Jinan, Shandong Province, China*

[2]*Shandong Normal University, Jinan, Shandong Province, China*

Abstract

The paper proposes the Dynamic Feedforward Neural Network based on Hidden Particle Swarm Optimization (HPSO-DFNN) to deal with the model predicative control problem of unknown nonlinear delay systems. It realizes quick, precise system modelling for controlled objects. Besides, the Smith predictive double controllers are designed to separate fixed set point control from external disturbance. The DFNN based on large-scale PSO is treated as an identifier and a predictor for the complex controlled objects with the purpose of increasing the robustness of the control system. Furthermore, aiming at the problem of constrained multi-input-multi-output (MIMO) model predictive control, rolling optimization is conducted to obtain controlled quantity through the PSO algorithm. After that, a combined neural network structure is put forward and applied to system modelling. Finally, the paper uses the typical nonlinear model to verify its effectiveness.

Keywords: nonlinear delay systems, dynamic feedforward neural network, particle swarm optimization

1 Introduction

Neural networks (NNs) are divided into two types in terms of network structure: recurrent neural network and feed forward neural network (FFNN). Compared with the former, the latter has advantages such as simple structure, fast calculation and distinctive layers. However, FFNN falls under the category of static networks. Moreover, it has poor generalization ability in dynamic systems. Among all learning algorithms, the paper chooses the general gradient descent algorithm for repeated reverse derivation. But it is easy to fall into local minima. Moreover, the adjustment formulas of parameters differ from each other [1]. A good deal of formula derivation is thus indispensable. Obviously, it requires excellent mathematics. The birth of PSO provides access to the problem. Nonetheless, the simple combination of particle swarms and NNs does not suffice to improve accuracy effectively. This is because the PSO has a "premature" problem. In the neural network training process, it often abandons searching ahead of schedule. Moreover, the ability of continuous approximation plays a more important role during the latter part of the process of NN identification. Therefore, it is of great significance to design a simple, feasible method, which can improve the efficiency of NN modelling with higher calculation speed and accuracy [2].

In addition, there is usually a lot of interference in practical control systems. The Smith predictive double controllers are designed to distinguish fixed set point control from disturbance control and enhance the robustness of the entire system. As for more complicated controlled objects, the FT-HPSO algorithm is chosen in the control system to optimize FFNN parameters and solve the large-scale optimization problem. The FT-HPSO-DFNN, as a predictor and an identifier, helps to simulate the predicted output value and the current output value respectively of the controlled object.

The above research is generally limited to unrestrained single/multiple input and single output control systems. In the paper, FT-HPSO-DFNN is used for unknown controlled object modelling, which conforms to the actual situation better and helps solve the constrained MIMO problem. Furthermore, rolling optimization is carried out and the controlled quantity can be figured out. Then, the paper puts forward the model predictive control structure based on combinatorial optimization. The effectiveness of the structure is then verified through a simulation experiment on the predictive control of the multivariable nonlinear controlled objects.

2 PSO algorithm design of neural network training

BP neural network technology is a mature neural network approach, which has been applied in various fields. However, there still exit some corresponding shortages, which mainly include that: easy to fall into local minima; unable to accurately determine the number of hidden layer nodes of the network; slow convergence and so on. In recent years, many scholars have proposed a variety of methods to solve the above problems while others dedicated to using other optimized algorithm to train

---

* *Corresponding author* e-mail: 20057789@sdufe.edu.cn

neural network. But in the existing algorithms, the network's hidden layer nodes are still difficult to be determined. Therefore, this paper attempts to propose a HPSO-NN with hidden layer nodes, aiming to set the number of nodes in hidden layer in BP network as an important optimization goal for particle swarm algorithm, and set the numbers of the hidden nodes and the each weight and threshold value of BP network jointly as an important optimization goal for particle swarm algorithm [3].

As a simple, effective stochastic search algorithm, PSO algorithm could be used to optimize the neural network. Although research in this regard is still in its infant stage, some research results have demonstrated that the PSO algorithm has a great potential in terms of optimization of neural networks.

The utilization of PSO algorithm as the study algorithm of neural network mainly due to fact that the search of PSO algorithm does not rely on gradient information and it simply requires the feasible solution of the function under the constraints as well as the ability of global search. Using the PSO algorithm could conduct a global research on the weight and threshold value of neural network as well as the hidden layer nodes until it meets the demand of accuracy [4]. For PSO algorithm is based on population, and it can get many extreme values from different nodes, this algorithm could not be easily to fall into local optimal solution, which can solve the problems of BP neural network effectively and improve the generalization ability of neural networks.

2.1 THE DESCRIPTION OF THE SOLVED PROBLEMS

Using PSO algorithm to train the neural network, for a given BP neural network shown in Figure 1, the parameters need to be optimized are: the number of the hidden layer nodes in network, the connection weights w_{ji} from the input layer, the threshold value b_{1j} in hidden layer, and the threshold value b_{2k} in output layer. For these parameters, using the corresponding encoding method to encode them could create a single particle, which is the vector space solution to solve the problem. The position of each particle is the corresponding solution, and its speed is the basis of the adjustment of the position. Many particles could form a particle groups. The iterative process of the algorithm requires a fitness function. By calculating the degree of adaptation, the quality of the particles could be evaluated to determine the optimal solution globally in each generation and the individual optimal solution. Each particle has its own speed and position [5]. Through a global optimal solution and the optimal solution as well as individual specific formula, the particle velocity and position could be adjusted to make it tend to the optimal solution. When the iteration error is less than a certain algebraic ε or reaches a given maximum number, the calculation could be stopped to find the optimal solution. Then the optimal solution could be turned into the weight

and threshold value of BP neural network and a good optimized neural network could be obtained. This is the process to solve the problem [6].
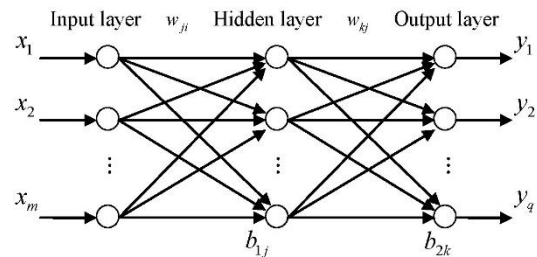
FIGURE 1 BP neural network

2.2 THE ENCODING METHOD OF THE PARTICLES

According to the neural network structure shown in Figure 1, the number of the network input and output layer nodes are m and q. Assuming that the maximum number of hidden nodes are n, then the number of connection weight w_{ji} from output layer to hidden layer are $m \times n$, the number of w_{kj} are $n \times q$, the threshold value b_{1j} are n, b_{2k} are q. So, the number of parameters wait to be optimized in BP neural network are total $1+m \times n+n \times q+n+q$, among which 1 refers to be the number of hidden layer nodes await to be optimized. In all, the PSO algorithm is to find the optimal number of hidden nodes and the corresponding weights and thresholds in this dimensional space.

Depending on the encoding objects, each particle is divided into two parts, the head and body, in which the body can be divided into four parts which could be expressed graphically in Figure 2:
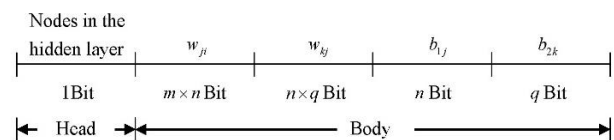
FIGURE 2 Structure of particles

There is only 1 part in the region of head, consisting of the information of the number of the hidden layer nodes in BP network. As the number of hidden layer nodes must be an integer, the head part could be taken the measure of a pseudo-integer encoding, that is, to use real number encoding method in the encoding process, and get the round when calculating the fitness of particles. Thus, an integer of information could be achieved. However, the position and velocity of a particle used in the equation is still a true real number [7].

The ranges of the real numbers in the head part of the particles are determined by the number of the hidden layer nodes in the scope of BP neural network. When conducting the initialization of the particles, the initialization of the real numbers in the head part could be determined by the possible scope of the hidden layer nodes. If the maximum

number of nodes in the hidden layer of BP neural network is C_{max}, the minimum could be C_{min}, thus the number of hidden layer nodes can take $C_{max} - C_{min} + 1$. So the ranges of the real number c in the head of particles are $[C_{min} - 0.5, C_{max} + 0.5)$. The reason for adding (minus) 0.5 on both digital range, as it is the only way to make the head portion of particles with equal probability indicates the number of hidden layer nodes. Real number of particles through the head store will be rounded to the nearest whole number obtained in the hidden layer nodes corresponding particle BP neural network.

For the convenience of the following description, the performance code is defined like this:

Performance Code: the real number obtained by integer rounding stored in the head portion is defined as the performance code of the particle. The performance code is the number of particles in the hidden layer node that it represents in BP neural network [8].

The performance code is very important. It shows the number of hidden layer nodes of the current particle under its corresponding BP neural network and plays a guidance role for further fitness calculation of particle and adjustment of particle position and velocity. The performance code is also an effective tool for the researchers to observe the number of the hidden layer nodes during the movement process of particle. Different particles have different performance code, along with different number of hidden layer nodes. Thus, it enables the particles to find different number of hidden layer nodes in space, as well as find out the most proper number of hidden layer nodes in constant search and adjustment of position and speed. By doing so, the structure of BP network could be determined to reach the optimization of the number of hidden layer nodes.

2.2.1 The part of head

Although there is only one real number in this region, it plays a big role. It is equivalent to the soul of particles, and all future operations on the particles depend on the special consideration of this part.

2.2.2 The part of body

Since all weight and threshold values are real numbers, the body part uses the method of real number encoding. Depending on the encoding object, the body part is divided into four smaller parts, representing the connection weight w_{ji} from input layer to hidden layer in BP neural network, the connection weight w_{kj} from hidden layer to output layer, the threshold value b_{1y} in hidden layer as well as b_{2k} in output layer.

The encoding of the body region is designed according to the maximum number of hidden layer nodes (n), to fully reflect all the hidden layer nodes in the head part, in case that there is no enough digits for the body part to save the weight and threshold values. As the performance code is

different, it is impossible to reflect the maximum possible number of the hidden layer nodes by performance code, which will inevitably lead to the situation that when the performance code is not the maximum possible number of hidden layer nodes, some rear part of the body could not be used and the calculation in respect of the bit is invalid [9].

3 Particle swarm neural network algorithm with optimized hidden layer nodes

3.1 THE INITIALIZATION OF PARTICLE SWARM

According the way of encoding, before the beginning of the algorithm, you need to initialize the particle swarm work. In this process, you need to form a population based on population size and the numbers of random individuals (particles), in which different individuals represent different neural network structure and different set of weight and threshold value. The speed of each particle and g_{best} and several p_{best} also need to be initialized.

3.2 DESIGN OF NEURAL NETWORK TRAINING AND FITNESS FUNCTION

The optimization of network weight value is an iterative process. Normally in order to ensure that the training neural network has strong generalization ability, in the process of training the network, the given sample space is divided into two parts: one part is the training sample, called a training set; another part is a test sample, called a test set. With the training set to train the network, using the test set to evaluate the accuracy and generalization ability of the network [10].

In order to determine the individual (particle) the merits of the degree in the population, a uniform standard to measure is needed. The following definitions of particle's fitness are given like this:

Fitness: a standard population used to assess the extent of the pros and cons of each particle. Fitness particles usually calculated by the fitness function. Depending on the specific problem, the design of the fitness function is different.

The design of the fitness function is to assess the extent of the major merits of each particle, which is the driving force of evolution. The degree of the merits of the particles is that it represents the accuracy of BP neural network. The smaller the network error on the smaller sample of learning is, the better the fitness of particles is. Hence, the fitness function could be defined like this:

$$Fitness = \frac{1}{q} \sum_{p=1}^{N} \sum_{k=1}^{q} \left(y_{pk} - o_{pk} \right)^2 , \qquad (1)$$

where N is the number of training samples, q is the number of output layer nodes in the BP neural network, y_{pk}, o_{pk} are actual output and the target output of the p samples at the k output node. As it can be seen from the fitness function,

the smaller the fitness value of particles is, the smaller of the error of BP network is, and the smaller the particles are.

BP neural network training process is: calculated based on particle swarm in the first part of the performance of each individual code to determine the number of hidden layer neural network nodes, and then they are part of the body is mapped to the network weights and thresholds, which constitutes a neural network [11]. Corresponding to each individual neural network input training samples for training. When finished entering all of the samples to calculate the fitness of particles, provide the basis for adjusting the speed and position of the particles, the particles closer to the global final solution to achieve the training of the network.

3.3 THE TERMINATION CONDITION OF THE ALGORITHM OF THE PARTICLE VELOCITY AND THE ADJUSTMENT OF POSITION

The adjustment of the particle velocity is mainly based on the population of the globally optimal solution and its own individual optimal solution. It is conventional to adjust the position of the superimposed position and its speed. And the adjustment of the position of the particle velocity at time $t + 1$ is shown in the following formula [12]:

$$v(t+1) = wv(t) + c_1 r_1 \left(p_{best}(t) - x(t) \right) + c_2 r_2 \left(g_{best}(t) - x(t) \right), \tag{2}$$

$$x(t+1) = x(t) + v(t+1), \tag{3}$$

where w is the inertia factor, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are called to the interval [0,1] of two independent uniformly distributed random numbers. Parameter w, c_1 and c_2 values depend on the specific issues.

The conditions for the termination of algorithm:

Condition 1: when the fitness of the particle is less than the given $\varepsilon(\varepsilon \to 0)$, the algorithm is terminated.

Condition 2: when the revolution of the population reaches to some given upper limit T, the algorithm is terminated.

Condition 3: when the gap of the optimized fitness of the population is less than ε, the algorithm is terminated.

Conditions 1 and 2 is the original PSO algorithm to optimize the conditions for BP neural network that already exists on the basis of this paper, in order to overcome the network. Over-fitting, phenomenon proposed a third termination condition. When the N consecutive optimal fitness groups are less than ε, a majority of the particles has evolved well. If this algorithm continues, although the network will reduce the error, the generalization capability of the network will gradually decrease.

When the algorithm terminates, put out the global optimal solution. The first part of the calculated performance code could be as BP neural network hidden layer nodes, and it will be mapped as part of the body weights and thresholds of the network. Finally we get the optimized BP neural network [13].

3.4 THE ALGORITHM FLOW CHART

The main purpose of hidden layer nodes with the purpose of optimizing the particle swarm neural network algorithm is to seek the number of nodes and weight and threshold value that most closely matches the initial neural network hidden layer. For different initialization, it will come to different hidden layer junction point number. Firstly, we should set the required parameters, and then according to the number and scope of the position and velocity of the particle particles population, we could randomly generate particle population. For each generation of particles, the particles must first be calculated on performance code that represents the number of hidden layer nodes in BP neural network, and then map into the particle body parts neural network weight and threshold value, and also calculate the fitness of particles degrees, and further determine the global optimum and individual optimal solution. We finally generate a new generation of particles continue iteration. Until the fitness particles are less than a given number of iterations or meet the requirements, the algorithm terminates.
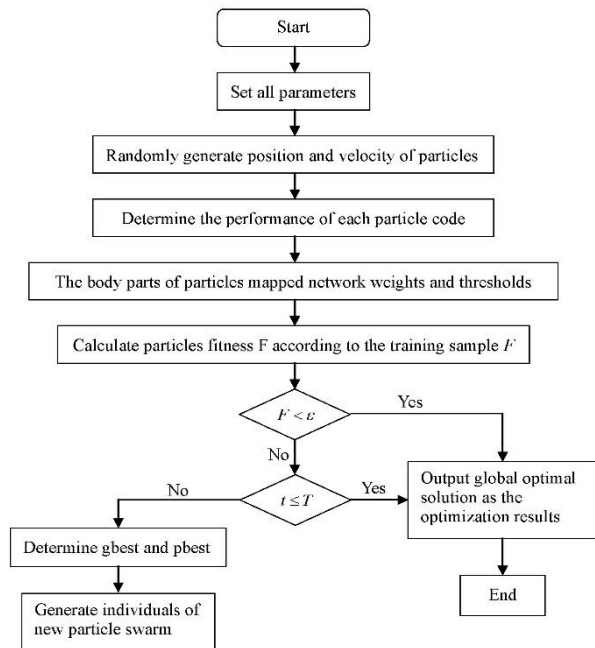
FIGURE 3 Flow chart of the algorithm

4 Smith predictive double controllers based on cooperative large-scale PSO

The paper first uses FT-HPSO to optimize the proposed DFNN parameters and thus fulfil the FT-HPSO-DFNN model. The FT-HPSO large-scale optimization algorithm is adopted in the face of more complicated controlled objects. It is necessary to choose the NN structure that

contains more hidden layers so as to increase network information and approximation capability. Take the proposed "2×20×20×1" DFNN for example. The dimension of the parameter to be optimized in the model is 561. Ordinary PSO algorithms cannot realize excellent optimization. The FT- HPSO-DFNN model plays an important role in Smith double controllers in two ways. As an identifier, it can predict the output of the controlled object and correct modelling errors. As a predictor, it is able to predict the output of the deferred object. After removing the dynamic delay operators in DFNN output layer, the delay time in the system can be eliminated and the change in controlled objects can be predicted ahead of time. Furthermore, the proposed double controllers can separate fixed set point control from interference rejection and enhance the system robustness.

4.1 DESIGN OF SMITH PREDICTIVE DOUBLE CONTROLLERS

The Smith predictive double controllers based on FT-HPSO-DFNN has the predictive control structure as shown in Figure 4:
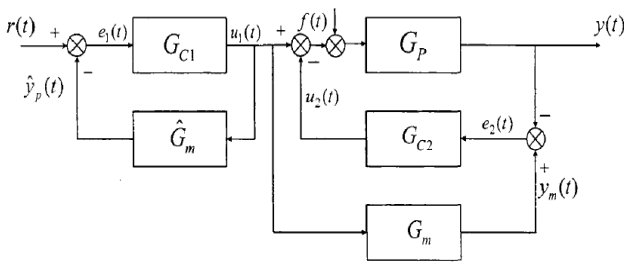
FIGURE 4 Structure of Smith predictive double controllers system

As shown in the above figure, the system structure includes two controllers, G_{C1} and G_{C2}, which give rise to two controlled quantity, $u_1(t)$ and $u_2(t)$. The former plays a part in controlled object G_p, identifier G_m, and predictor $\hat{G}_m$. System error $e_1(t)$ is the difference between set value $r(t)$ and predicated output $\hat{y}_p(t)$. $e_2(t)$ is the difference between the output of identifier G_m and controlled object G_p. In an ideal condition when $G_m=G_p$, $e_2(t)=0$, and the output of controlled object G_p is determined by the output $u_1(t)$ of controller G_{C1}. The system is a typical Smith predictive control structure. On account of the interference of $f(t)$ or the error in modelling process, $G_m \neq G$ and deviation inevitably occurs between G_p and $r(t)$ under the control of G_{C1}. The model error compensation controller G_{C2} is introduced to get rid of the situations such as $e_2(t) \neq 0$. The proposed Smith predicative double controllers have better robustness owning to a full consideration of the error in modelling process as well as the influence of external disturbance.

The transfer function of the entire system is:

$$Y(s) = H_r(s)R(s) + H_f(s)F(s),\qquad(4)$$

where $H_r(s)$ and $H_f(s)$ are the transfer functions of fixed set point control and load disturbance, respectively. Below are their expressions:

$$H_r(s) = \frac{G_{C1}(s)P(s)e^{-ds}}{\left[1+G_{C1}(s)\hat{G}_m(s)\right]}\frac{\left[1+G_{C2}(s)\hat{G}_m(s)e^{-d*s}\right]}{\left[1+G_{C2}(s)P(s)e^{-ds}\right]},\qquad(5)$$

$$H_f(s) = \frac{G_P(s)e^{-ds}}{1+G_{C2}(s)G_P(s)e^{-ds}}.\qquad(6)$$

4.2 CONTROL STRUCTURE & PERFORMANCE ANALYSIS

The normal PI controller is adopted as shown in Equation (7) for better performance analysis of the controller.

$$G_C(s) = K_C\left(1+\frac{1}{T_C s}\right).\qquad(7)$$

Describe the controlled object G_P by a first-order inertial element in Equation (8). So the identification model G_m is as shown in Equation (9).

$$G_P(s) = P(s)e^{-ds} = \frac{K_P}{T_P s+1}e^{-ds},\qquad(8)$$

$$G_m = \hat{G}_m e^{-d*s} = K_P^*\left(1+\frac{1}{T_P^* s}\right)e^{-d*s},\qquad(9)$$

where d is delay time and d^* is delay time parameter after identification. Put them into the equation, and the results are:

$$H_r(s) = \frac{K_{C1}K_P(T_{C1}s+1)e^{-ds}}{T_{C1}s(T_P^* s+1)+K_{C1}K_P^*(T_{C1}s+1)} \cdot$$
$$\frac{T_{C2}s(T_P^* s+1)+K_{C2}K_P^*(T_{C2}s+1)e^{-d*s}}{T_{C2}s(T_P s+1)+K_{C2}K_P(T_{C2}s+1)e^{-ds}},\qquad(10)$$

$$H_f(s) = \frac{K_P T_{C2}se^{-ds}}{T_{C2}s(T_P s+1)+K_{C2}K_P(T_{C2}s+1)^{-ds}}.\qquad(11)$$

In ideal conditions, $G_p(s)=G_m(s)$ and fixed set point control $H_r(s) = \frac{G_{C1}(s)G_P(s)e^{-ds}}{1+G_{C1}(s)G_P(s)}$. It is only determined by fixed set point controller $G_{C1}(s)$, while $H_f(s)$ is only determined by $G_{C2}(s)$. Since $\lim_{s\to 0}H_r(s)=1$ and $\lim_{s\to 0}H_f(s)=0$, there are no steady-state errors in the system. Output $y(t)$ can also follow the change of set value $r(t)$ even in dual-ring, contact-free structures. The control signal:

$$U_1(s) = \frac{G_{C1}(s)}{1+G_{C1}(s)\hat{G}_m(s)}R(s) =$$

$$\frac{K_{C1}(T_P^*s+1)(T_{C1}s+1)R(s)}{T_{C1}s(T_P^*s+1)+K_{C1}K_P^*((T_{C1}s+1)}, \qquad (12)$$

So $\lim_{s\to0}U_1(s) = \frac{1}{K_P}R$. Similarly, $\lim_{s\to0}U_2(s) = F$. The total controlled quantity of the control system is:

$$\lim_{s\to0}U(s) = \lim_{s\to0}(U_1(s)-U_2(s)) = \frac{1}{K_P}R - F .$$

In Smith control structure, model precision produces a great influence on control effect. The insensitivity of model mismatch can be analysed when $K_P^* \neq K_P, T_P^* \neq T_P$ and $d^* \neq d$. The controlled quantity $\lim_{s\to0}U_1(s) = \frac{1}{K_P^*}R$, while $\lim_{s\to0}U_2(s) = F + (\frac{1}{K_P^*}-\frac{1}{K_P})R$. Thus, the total controlled quantity remains the same:

$$\lim_{s\to0}U(s) = \lim_{s\to0}(U_1(s)-U_2(s)) = \frac{1}{K_P}R - F . \qquad (13)$$

From above we can draw the following conclusions. Model mismatch has no influence on control signals but enhances system robustness. $G_{C1}(s)$ and $G_{C2}(s)$ are mainly influenced by fixed set point controller and barrage jamming. The deviation resulted from model mismatch can be eliminated through $G_{C1}(s)$ and $G_{C2}(s)$.

4.3 DESIGN OF NEURAL NETWORK CONTROLLER

As for controller $G_{C1}(s)$, a three-layer BP neural network is adopted. The controlled quantity is the NN output, which is shown as follows:

$$u_1(t) = \sum_{i=1}^{n}\omega_2(t)\text{sig}\left(\sum_{j=1}^{m}\omega_1(t)e_1(t)\right), \qquad (14)$$

where $e_1(t) = r(t) - y_p(t)$; ω_1 and ω_2 are respectively the connection weight of the network; m and n represent the neuron numbers at the input level and the hidden level. The objective function is defined as $J_1(t) = \frac{1}{2}e_1^2(t)$ and then the weight can be updated as the following expression shows:

$$\omega(t+1) = \omega(t) - \eta_1\frac{\partial J(t)}{\partial\omega(t)}, \qquad (15)$$

where η_1 is the learning rate of controller $G_{C1}(s)$ and $\frac{\partial J(t)}{\partial\omega(t)} = -e_1(t)\frac{\partial y_p(t)}{\partial\omega(t)}$. So Equation (15) becomes:

$$\omega(t+1) = \omega(t) - \eta_1 e_1(t)\frac{\partial y_p(t)}{\partial\omega(t)} . \qquad (16)$$

Similarly, the control law of controller $G_{C1}(s)$ is as Equation (17) shows:

$$u_2(t+1) = u_2(t) + \eta_2 e_2(t)\frac{\partial y(t)}{\partial u_2(t)}, \qquad (17)$$

where η_2 is the learning rate of controller $G_{C2}(s)$.

5 Model predictive control of unknown nonlinear dynamic systems

The above model predictive control helps obtain the future output on the basis of controlled object equation. But it is necessary to use NN for system identification and multistep prediction when the controlled object is a model unknown nonlinear dynamic system. A combined NN model based on FT-HPSO-DFNN is used for system identification. The model parameter built in accordance with the NN after system modelling is as follows:

$$y(k) = g(u(k-d), u(k-d-1),...,u(k-d-m),$$
$$y(k-1), y(k-2),..., y(k-n)), \qquad (18)$$

where $g(\cdot,\cdot)$ stands for a nonlinear function. d, u, and y represent system delay, input and output of the controlled object, respectively. The DFNN structure mentioned is then adopted. After off-line system identification, the dynamic delay operators between output layer and hidden layer are viewed as the input, while the variables in the hidden layer are expressed as the output. Definitions:

$$\varphi_m(k) = [u(k-d), u(k-d-1),...,u(k-d-m)]^T$$

and

$$\varphi_n = [y(k-1), y(k-0),..., y(k-n)]^T .$$

The single-step predicated output of the NN is:

$$\hat{y}(k|\theta(k)) = F\left(\sum_{j=1}^{n_h}\omega_{i,j}f_j(a) + b_{i,0}\right), \qquad (19)$$

$$a = \sum_{l_m=1}^{n_m}\omega_{j,l_m}\varphi_{l_m}(k) + \sum_{l_n=1}^{n_n}\omega_{j,l_n}\varphi_{l_n}(k-1) + b_{j,0}, \quad (20)$$

where j is the neural network number in the hidden layer; ω_{j,l_m} , ω_{j,l_n} and $\omega_{i,j}$ are the threshold values of the hidden layer and the output layer; $b_{j,0}$ and $b_{i,0}$ mean the bias of the hidden layer and the output layer; f_j and F_i are the excitation functions of the hidden layer and the output layer; $\theta(k)$ represents all adjustable parameters at the moment k.

So multistep output prediction of the controlled object is realized the combined neural network model based on

FT-HPSO-DFNN. Multistep prediction is carried out by dint of the neural network. It is divided into two categories: recursive multistep prediction and non-recursive multistep prediction. The former uses one-step prediction and obtains η-step ahead prediction through repeated equation solving. The latter predicts the $k + \eta$ step by directly using the input and output data at and before the moment k. Thus, the two have different nonlinear functions because their predicted horizons are not the same. Consequently, calculation becomes more complicated. Therefore, the paper adopts the recursive multistep prediction method. Compared with Equation (18), the η-step ahead prediction is figured out:

$$\hat{y}(k + \eta | k) =$$
$$g(u(k+\eta-d), u(k+\eta-d-1), ..., u(k+\eta-d-m)$$
$$y(k-1), y(k-2), ..., y(k - \max(n-\eta, 0)), \qquad (21)$$
$$\hat{y}(k+\eta-1), \hat{y}(k+\eta-\min(\eta, n)))$$

$\hat{y}(k + \eta | k)$ is the output of controlled object at the moment $k + \eta$, which is predicated at the moment k. Thus, the multistep predicated output of the NN is:

$$\hat{y}(k+\eta|k) = F_i(\sum_{j=1}^{n_h} \omega_{i,j} f_j(\tilde{a}(\eta, j)) + b_{i,0}), \qquad (22)$$

$$\tilde{a}(\eta, j) = \sum_{\lambda=1}^{\min(\eta-d,n)} \omega_{j,\lambda} \hat{y}(k+\eta-\lambda) +$$
$$\sum_{\lambda=\min(\eta,n)+1}^{n} \omega_{j,\lambda} y(k+\eta-\lambda) + \qquad . \qquad (23)$$
$$\sum_{\lambda=0}^{m} \omega_{j,n+1+\lambda} u(k+\eta-d-\lambda) + b_{j,0}.$$

So the model predictive control problem is based on the proposed PSO so as to carry out constrained optimization and get the optimal controlled quantity.

The entire predictive control process of the PSO-based neural network is as follows:

Step 1: Make k=1, predicated horizon be N, control horizon be N_u, and weight matrixes be Q and R;

Step 2: Generate input and output data of the controlled object, and conduct off-line training for FT- HPSO-DFNN;

Step 3: After the NN model stabilizes, calculate predicated the output value $\hat{y}(k)$ of one-step controlled object using the well-trained NN at the sampling time k;

Step 4: At the time $k + N_u - 1$, obtain a series of control input $u(k + N_u - 1 | k)$ with the help of the MPC control strategy. And then get the predicated value $\hat{y}(k+\eta|k)$ of controlled object through the recursive model;

Step 5: Minimize the objective function using the PSO algorithm and work out the control increment $\Delta u(k)$;

Step 6: Figure out the control input $\Delta u(k) = \Delta u(k|k) + u(k-1)$;

Step 7: If k=k+1, return to Step 4.

6 Simulation analysis of model predictive control of multivariable nonlinear systems

The paper uses the multivariable nonlinear ball and plate system to verify the proposed research on model predictive control based on PSO and neural network system identification. The ball and plate system, as a multivariable nonlinear controlled object, is the two-dimensional extension of the ball and beam system. Its controlled object is a plate with two rings pivoted on axes at right angles to each other. The visual feedback enables the ball which rolls freely to maintain balance at any given position on the plate or roll along a particular track.

The state variable is defined as:

$$X = (x_1, x_2, ..., x_8)^T = (x, \dot{x}, \theta_x, \dot{\theta}_x, y, \dot{y}, \theta_y, \dot{\theta}_y)^T, \qquad (24)$$

where x and y are the positions of the ball. θ_x and θ_y are the rotor angles on the X axis and the Y axis. Conduct dynamic modelling for the controlled object of two-degree of freedom according to Lagrange differential equation. The state equation of the ball and plate system is:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{x}_7 \\ \dot{x}_8 \end{bmatrix} = \begin{bmatrix} x_2 \\ B(x_1 x_4^2 + x_4 x_5 x_8 - g(\sin x_3 - \mu \cos x_3)) \\ x_4 \\ 0 \\ x_3 \\ B(x_5 x_8^2 + x_1 x_4 x_8 - g(\sin x_7 - \mu \cos x_7)) \\ x_8 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_x \\ u_y \end{bmatrix}, \qquad (25)$$

where parameter B is defined as $B = \dfrac{m}{m + \dfrac{J_b}{R^2}}$. The ball

mass $m=0.1$kg, and the ball radius $R=0.02$m. The rotational inertia $J_b = 1.76\mathrm{e}^{-5}$Kg.m^2 . The input-output equation of the entire ball and plate system is expressed as:

$$[x, y]^{\mathrm{T}} = g(\theta_x, \theta_y) . \qquad (26)$$

Hence, the ball and plate system is a two-input, two-output nonlinear dynamic system that has eight state variables. Firstly, FT-HPSO-DFNN model is used for off-line system modelling. After that, the paper analyses the result of multistep prediction. Control the ball and plate system by dint of the PID controller so as to obtain 2,000 sets of input and output data. The former 1,000 sets of data are used for NN training and the rest for testing. Table 1 contains the predictive results of different NN model tests.

TABLE 1 The comparison of different NN structures predictive results

NN Structure	1-Step Prediction	5-Step Prediction	10-Step Prediction	100-Step Prediction
FT-HPSO-DFNN	1.33E-4	7.01E-4	7.12E-4	7.94E-4
ELM	5.40E-3	9.10E-3	1.43E-2	1.21E-1
ESN	5.61E-3	1.56E-2	2.29E-2	3.07E-2

The paper chooses three NN models: FT-HPSO-DFNN, ESN, and ELM to conduct multistep prediction and simulation. The result shows that precision declines as the predicated step length increases. As the FT-HPSO-DFNN model has high precision in the above 1-Step Prediction, the precision decreases slowly as the predicated step length increases. But the ELM algorithm leads to a quick rate of descent. Therefore, the proposed NN model can be well applied to unknown model predictive control of controlled objects. The precision remains quite high even in the situation of multistep prediction. Figure 5 is the fitted curve of 5-Step Prediction.

FIGURE 5 The 5 step predictive results of the proposed NN model

As shown in the above figure, the proposed FT-HPSO-DFNN model can realize multistep prediction well in the case of unknown models.

7 Conclusion

To integrate with practical control process better, the paper proposes a nonlinear model predictive control method based on cooperative large-scale PSO aiming at the problem of constrained MIMO model predictive control. Using the FT-HPSO-DFNN model for system modeling, the research realizes multi-step prediction for controlled objects. Furthermore, it adopts the PSO algorithm to optimize constrained objective functions and acquire the optimal control law. Based on simulation experiments of the ball and plate system, the paper proves the proposed model has good control effect and then provides some effective solutions.

Acknowledgments

References

[1] Kuo T J, Hung S Y, Cheng W C 2014 Application of an optimization artificial immune network and particle swarm optimization-based fuzzy neural network to an RFID-based positioning system *Information Sciences* **262** 78-98
[2] Coban C 2014 Power level control of the TRIGA Mark-II research reactor using the multifeedback layer neural network and the particle swarm optimization *Annals of Nuclear Energy* **69** 260-6
[3] Hajihassani M, Jahed Armaghani D, Sohaei H, Tonnizam Mohamad E, Marto A 2014 Prediction of airblast-overpressure induced by blasting using a hybrid artificial neural network and particle swarm optimization, *Applied Acoustics* **80** 57-67
[4] Khajeh M, Kaykhaii M, Hossein Hashemi S, Shakeri M 2014 Particle swarm optimization–artificial neural network modeling and optimization of leachable zinc from flour samples by miniaturized homogenous liquid–liquid microextraction *Journal of Food Composition and Analysis* **33**(1) 32-8
[5] de Mingo López L F, Gómez Blas N, Arteta A 2012 The optimal combination: Grammatical swarm, particle swarm optimization and neural networks *Journal of Computational Science* **3**(1-2) 46-55

[6] Green II R C, Wang L, Alam M 2012 Training neural networks using Central Force Optimization and Particle Swarm Optimization: Insights and comparisons *Expert Systems with Applications* **39**(1) 555-63
[7] Tsekouras G E, Tsimikas J 2013 On training RBF neural networks using input–output fuzzy clustering and particle swarm optimization *Fuzzy Sets and Systems* **221** 65-89
[8] Mirjalili S, Mohd Hashim S Z, Moradian Sardroudi H 2012 Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm *Applied Mathematics and Computation* **218**(22) 11125-37
[9] Valdez F, Melin P, Castillo O 2014 Modular Neural Networks architecture optimization with a new nature inspired method using a fuzzy combination of Particle Swarm Optimization and Genetic Algorithms *Information Sciences* **270** 143-53
[10] Ren C, An N, Wang K, Li K, Hu B, Shang D 2014 Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting *Knowledge-Based Systems* **56** 226-39

Operation Research and Decision Making

[11] Chen H S, Leou J J 2012 Saliency-directed color image interpolation using artificial neural network and particle swarm optimization *Journal of Visual Communication and Image Representation* **23**(2) 343-58

[12] Lazzús J A 2013 Neural network-particle swarm modeling to predict thermal properties *Mathematical and Computer Modelling* **57**(9-10 2408-18

[13] Dehuri S, Roy R, Cho S-B, Ghosh A 2012 An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification *Journal of Systems and Software* **85**(6) 1333-45

Authors

Lizheng Liu, born in February, 1980, Jinan, China

Current position, grades: lecturer in Shandong University of Finance and Economics.
University studies: M.Sc. Degree in Computer Science (2005, Shandong University).
Scientific interest: O2O (online-to-offline) commerce, business intelligence and mobile e-commerce technology.
Publications: 2 books, 2 scientific papers.

Fangai Liu, born in 1962, Qingdao, China

Current position, grades: professor in Shandong Normal University.
University studies: Ph.D in Computer Science (2002, Chinese Academy of Sciences).
Scientific interests: parallel processing, network optimization, e-commerce and grid computing.
Publications: 8 books, over 80 refereed articles and papers in scientific journals.
Experience: head of 3 National Natural Science Funds projects and 5 scientific and technological projects of Shandong province.

Feng Yang, born in August, 1972 in Jinan, China

Current position, grades: associate professor in Shandong University of Finance and Economics.
University studies: Ph.D in Computer Science (2005, Beijing institute of technology). Post doctor in High Computing Lab of Tsinghua University.
Scientific interests: O2O (online-to-offline) commerce, mobile e-commerce technology and mobile cloud computing.
Publications: 15 scientific papers, 9 patents.

Development and research on remote online education information system based on Web

Min Liu*, Jing Cao, Yanru Xue, Yinghua Yao, Xuezuo Zhao

Hebei Normal University of Science & Technology, Hebei, 066004, China

Abstract

Remote education is a way of developing education-teaching activities, which transmits outstanding teaching resources to massive students scattered in different time and space by the Internet, satellite and other methods of communication technologies. The Internet is the most convenient way to participate in distance learning. This paper applies B/S three-layer architecture and ADO data access technology to complete the design of the remote online education information system based on Web. The system according to the actual business needs of remote education, carry on the design based on the role of application and division of the task, and it composed of modules of courseware on demand, online management, online examination, online exercise, teacher management, student management, etc.

Keywords: remote online education, information system, Internet

1 Introduction

With the development of the society, people's demand for knowledge become stronger and stronger, we need more convenient ways of learning. Therefore, more and more people choose the Internet as a way of obtaining information. Remote education [1-4] is an autonomous learning way, which emphasizes the importance of student-oriented and practices the separation of teaching and learning. It is a new education way, which carries on a large scale teaching activities to transmit outstanding teaching resources to massive remote students scattered in different time and space. Compared with any kind of education mode, remote education puts more emphasis on technicality and scale, at the same time, the Internet is the most convenient way to participate in distance learning. Because the students thought, the web can receive all kinds of education resources actively. Therefore, to establish a web-based powerful remote education information system with advanced technology is a must.

Based on B/S three layer architecture and ADO data access technology, this paper designed a remote online education information system based on Web, which realized the five main teaching functions of network classroom, classroom live, courseware on demand, intelligent test and performance management [5], and put forward a real-time interactive teaching mode with an organic combination of live classroom and text interactive.

2 Structure design of remote education information system

2.1 SYSTEM WORKING MODEL

The objects and their associations of the system platform are referred to as the basic elements. The analysis of the system-working model is to determine the relevance between objects and object properties, and relevance between objects themselves. Make clear of the relevance between the events and give the object model and function model in the platform, thus to grasp the key features of the platform. This system consists of client and server, the system objects are divided into three categories correspondingly:

1. Client object - browser: the main part of client operating entity with standard properties and methods.

Browser Extension: the optional part of client operating entity. With standard properties and methods when the subject type is determined.

2. Server object - Web server: the main part of server functional entity with standard properties and methods.

Web server extension: the optional part of server functional entity. With standard properties and methods when the subject type is determined. Concrete object of Web server extension applied in this system is ASP (Active Server pages) object.

Back server: the independent server functional entity with standard properties and methods. Mainly includes the database management server, file transfer server and streaming media server.

Gateway: the server functional entity interface with standard properties and methods.

Corresponding author e-mail: ygyao@qq.com

Agent: the server functional entity interface with standard properties and methods.

3. Protocol Packets: the transport entity of client and server with standard properties and methods. The correlation between the object is shown in Figure 1.
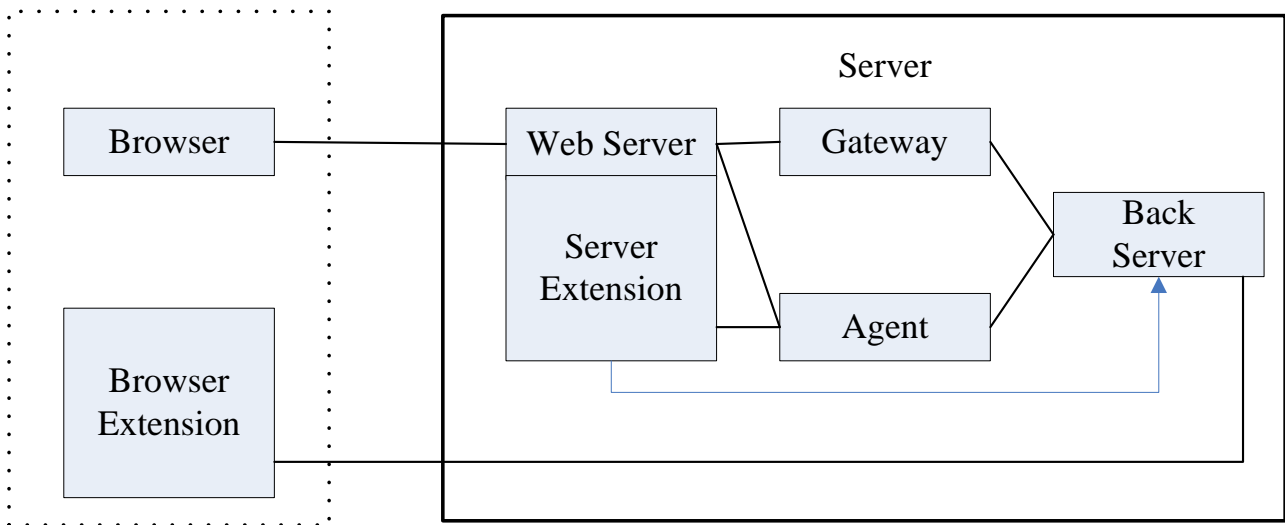
FIGURE 1 Object association diagram of system working model

According to different functions, the system is divided into several modules, and communication between modules is realized by interfaces. This design method can make the system structure more compact, the development more simple and clear and running more stable and reliable. System functional model diagram is shown in Figure 2.
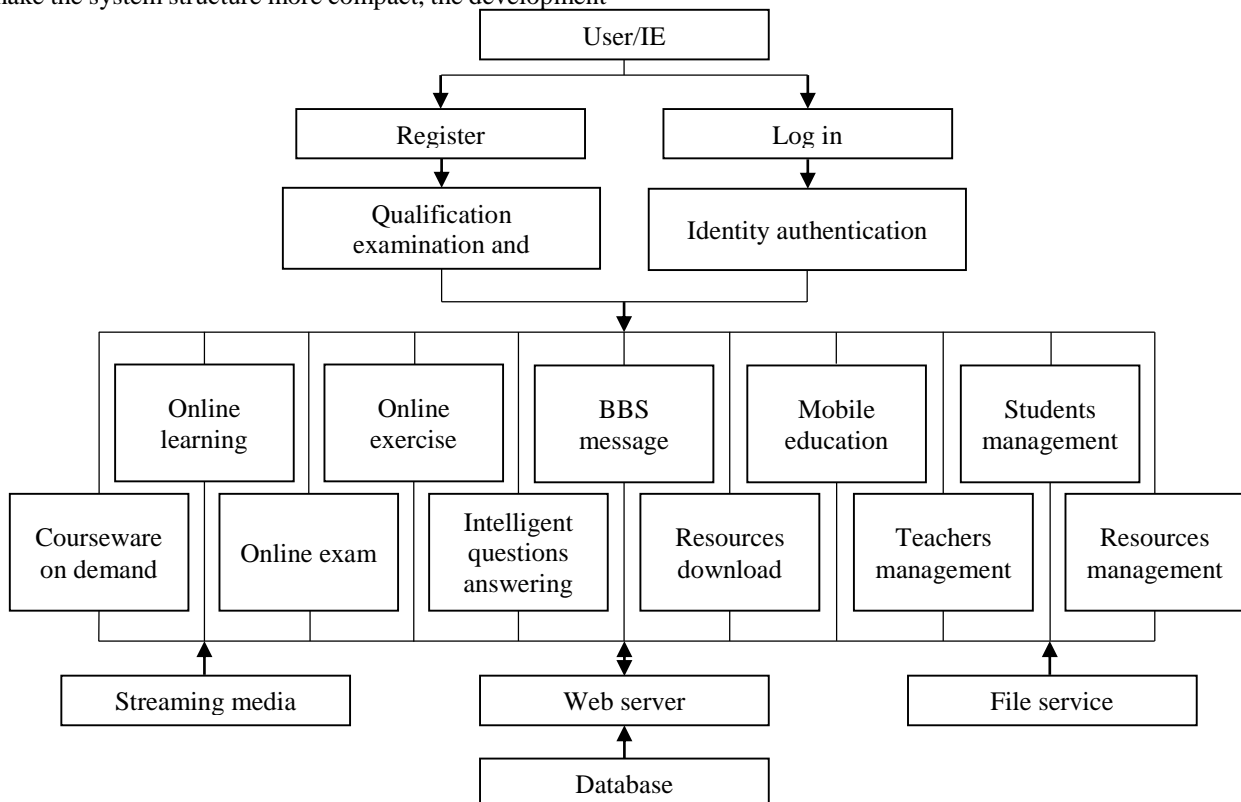
FIGURE 2 System functional model diagram

2.2 SYSTEM ARCHITECTURE DESIGN

B/S mode is a new MIS system platform model based on the Web technology. It decomposes traditional C/S mode server into a data server and one or more application servers (Web server), to make a three layers structure of client/server system. Three layer structures is in terms of the case of master-slave type C/S structure, as shown in Figure 3.

313
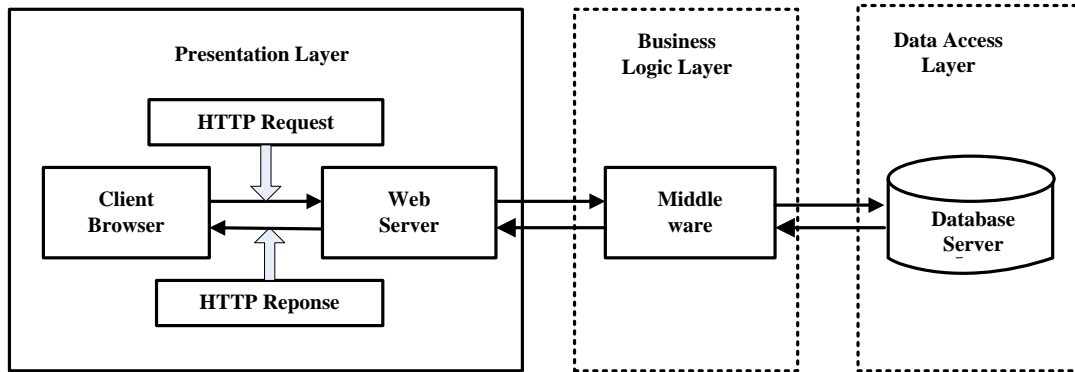
FIGURE 3 Three layer structure diagram

Three layer structure brings out the client's business logic part and put it in a running server specially, which is called business logic layer (be called as BLL for short). The layer near the user side is called presentation layer (be called as PL for short), and its responsibility is to assemble the data to a format. It not only includes the Web browser in the Web application, the Web server is also included [6-8], this layer provides information and data which can be displayed to the browser. The data access of database is called Data Access Layer (be called DAL for short). When a user submits the request, BLL layer is in response to user request and completes the transaction process. When it comes to database access, BLL layer needs to establish a connection with the database server for data access. As the connections in the database refers to connections established between BLL and database, and BLL layer connections can be shared, thus increases the number of users who can have access to the database.

B/S structure simplifies the client, and it does not need to install different client applications on different clients like C/S mode. Instead, it only needs to install the browser, which can simplify the development and maintenance of the system. System developers need not again to design and develop different client applications for different levels of users since all the functions are realized on a Web server. One more step we need to do is to set permissions for each group of users on different functions. It allows the user operation be more simple. For C/S mode, the client application has its own particular specifications and users

need to accept the special training. Whereas when adopting B/S mode, the client is just a simple and easy browser software. Both policy makers and operational personals can use it directly without any training. This characteristic of B/S mode leads to less limiting factors for platform management and maintenance. B/S is particularly suited to online information release and make the traditional MIS function expanded which cannot be achieved by C/S mode. And this new online information release function is needed for the modern distance education management, which makes the most of the distance education management paperwork be replaced by electronic documents, thus the efficiency of the remote management is improved. It also can play the advantages of distance education, simplify educational administration and other management procedures and save manpower and material resources.

Three layers structure has the easy expansion advantage. Due to the separation of the processing logic and webpage content, whatever changes of the business rules will not affect the front-end web content, and it is helpful for future maintenance. And the HTTP request will be through the firewall, which can block out many improper or malicious invasions.

This system adopts the three-layer structure based on B/S architecture. It is the server application system of execution. The client just needs to run standard browser software to complete all of the applications and functions. System object structure diagram is shown in Figure 4.
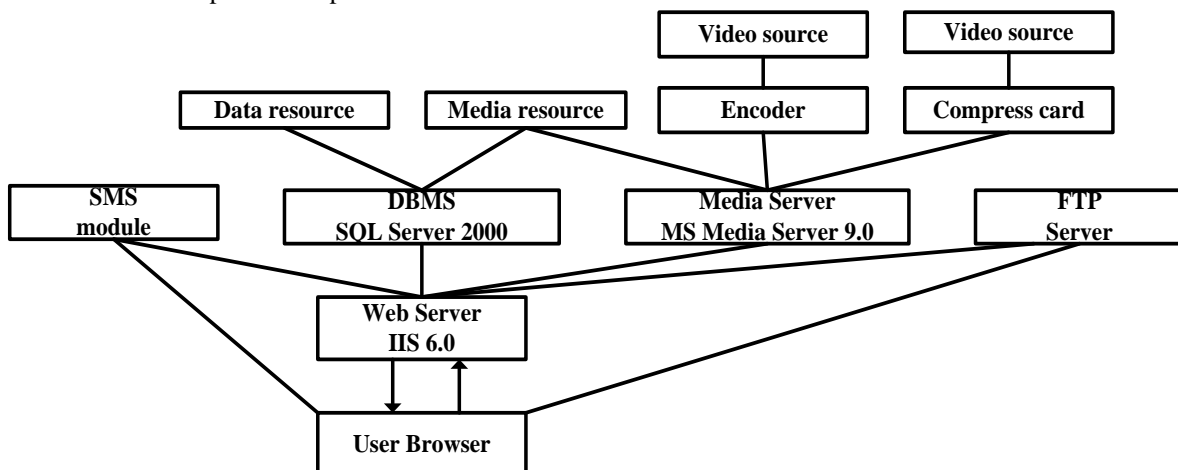
FIGURE 4 System object structure

3 Database design

3.1 THE DATABASE TABLE DESIGN

1) The users table is mainly used to save the user information with Id number automatically allocated by the

system so as not to be confused with other users. It includes the fields of username, login password, user's gender, age, phone number and corresponding user type. The primary key in the table is the Id number, the building model is shown in Table 1.

TABLE 1 The user table model

Field name	Description	Type	Length	Empty	Primary key
Id	User id	Varchar	25	No	Yes
Name	Username	Varchar	25	No	No
Password	User password	Varchar	20	No	No
Sex	User's gender	Varchar	10	Yes	No
Age	User's age	Int	4	Yes	No
Grade	User level	Varchar	20	Yes	No
Phone	User tel	Varchar	20	Yes	No
Type	User type	Varchar	10	Yes	No

2) CD-ROM resources information table. This table is used to store all the CD data in a database and convenient for management and operation of the administrator. It includes field such as CD serial number, CD version, CD

grade, CD discipline, CD type, CD name, and size of CD resources, upload time and other detailed information. The building model is shown in Figure 2.

TABLE 2 Resource information table

Field name	Description	Type	Length	Empty	Primary key
Id	Number	Varchar	25	No	Yes
Banben	Version	Varchar	500	No	No
Class	Grade	Varchar	100	No	No
Xueke	Discipline	Varchar	100	No	No
Type	Type	Varchar	100	No	No
Name	Name	Varchar	100	No	No
Daxiao	Size	Varchar	10	No	No
Date	Date	Datetime	8	No	No
Other	Remark	Varchar	500	Yes	No

3.2 JSP DATABASE CONNECTION TECHNOLOGY

1. Load JDBC driver: be sure to load driver before using the database. You can accomplish this task by applying for Name method of Class type, as follows:

Class. For Name ("Driver Name");

Driver Name is the name of JDBC driver to be loaded. The name can be determined according to the driver types provided by database vendors.

2. Establish a connection to the database: after loading the driver, establish a connection to the database, which needs to adopt the get Connection method of Driver Manager type. The form is generally as follows:

Connection conn=Driver Manager. Get Connection (URL, user, password)

The URL is a string representing data sources to be connected, i,e, the location of the database, the format is as follows:

jdbc: driver: database

3. Execute queries or other commands: as the database connection is established, you can use the SQL command. Before sending SQL commands, you need to create a Statement object, this object is responsible for the sending

of SQL statements to the database. After creating the Statement object, the SELECT query can be executed by execute Query method of the object. This method can return one Result Set object, which includes the query result. If you want to perform insert, update, delete command, then execute Update method of Statement object can be used to establish the Statement object syntax format as follows:

Statement sql=con.create Statement ()

Perform a variety of SQL statement syntax formats as follows:

*Result Set rs=sql. Execute Query ("select * from table")*

4. Handle the Result Set: Result Set object receives the result of executing the SELECT query. This object provides a variety of methods for accessing the database. Within each of the Result Set object, there is a record pointer. With the help of the movement of the pointer, you can traverse the data items of the object. Syntax format is as follows:

While (rs.next())
{
Rs.get String ("name");
Rs.get Int("age");
}

315

5. Close the connection: after the operation, you need to close the database connection in a timely manner to release the resources occupied. The order of closing objects is as follows:

1) Close Result Set, syntax format: rs.close();
2) Close Statement, syntax format: *sql.close();*
3) Close the connection, syntax format: *con.close().*

4 Implementation of remote education information system

4.1 OVERALL DESIGN OF SYSTEM FUNCTION

According to the standard reference model of remote education technology standard system of our country. A typical remote education system must contain three parts: the backbone system, management system and teaching resource database. The backbone system is composed of support system and supporting system. Support system is mainly responsible for the necessary functions in daily remote education teaching, which mainly refers to teacher instruction, and student's online learning functions,

including the network classroom, courseware on demand, online classes, mobile education, intelligent test and result management part. Supporting system aims to better cooperate with teaching, add a few auxiliary functions to meet the needs of individualized learning. It is an auxiliary system and plays a sidetrack role in remote education system, yet an indispensable part. Supporting system mainly includes news and announcement, online exercise, score inquiry, resources download and BBS part.

A complete and reliable system must be able to realize effective management and control for users and resources. Distance education system is no exception. It requires giving necessary control and monitoring of the process and quality of teaching and learning as well as user permissions and media resources. In order to realize reliable management of the system, the users are divided into teachers, students and administrators [9], each given different permissions and enter the system from three different paths (as shown in Figure 5). Different users with different permissions have different functions and operation interfaces to ensure the reliable operation of the system and normal and orderly conduct of online teaching.
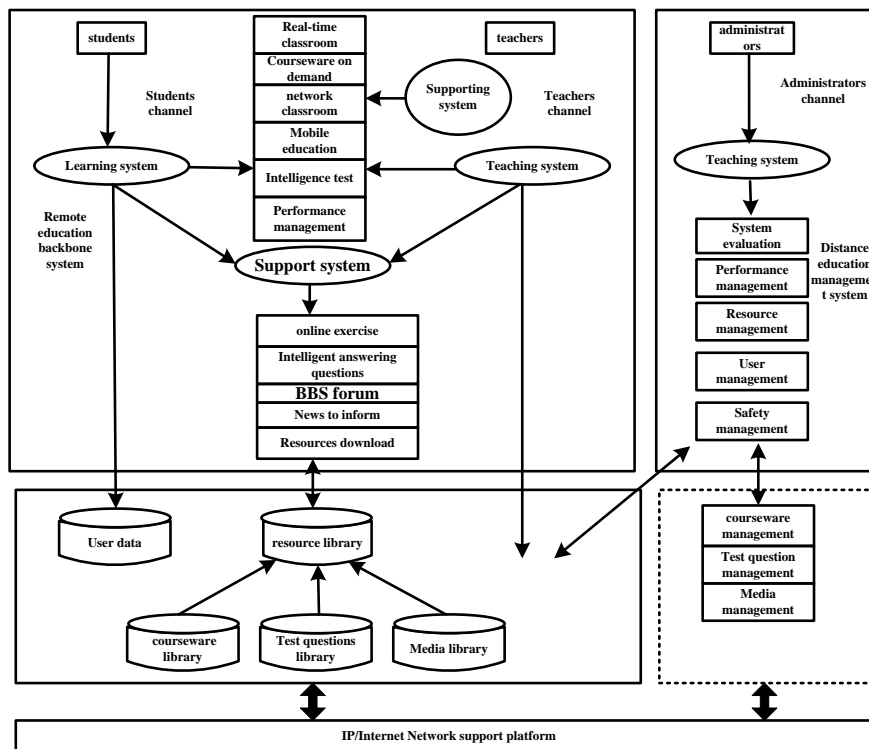
FIGURE 5 System function module chart

4.2 USER LOGIN MODULE DESIGN

The main function of the user login module is to distinguish the user identity permissions and validate whether user login password is correct or not. Therefore, this design provides three login channels for teachers, students and administrators, as shown in Figure 6. Teachers are the organizers and managers of teaching activities and teaching process of the whole teaching

system with their main task to host teaching activities. Teachers' authority is based on curriculum. They do not have any rights on the courses beyond their permissions. Students as legitimate users of the teaching system, before login must be registered in the remote management system, finish course selection and pay cost. Students have right to comprehensive learning and join a variety of online teaching activities on selected courses. Administrators'

responsibility is to perform management and maintenance of the teaching content in the whole teaching system.

In order to strengthen the security performance of the system, prevent the illegal user's logging in by usurpation of legitimate user account through history information on the computer; the system design adds a random authentication code. Only when all of the three of user login name (or ID, including teachers and student ID number), password and random authentication code are right can the user be authenticated and operations that match with the user permission be performed. The user password and random authentication code all use MD5 encryption and then be transferred from the client to the server [10], which further ensures the safety and reliability of the system.
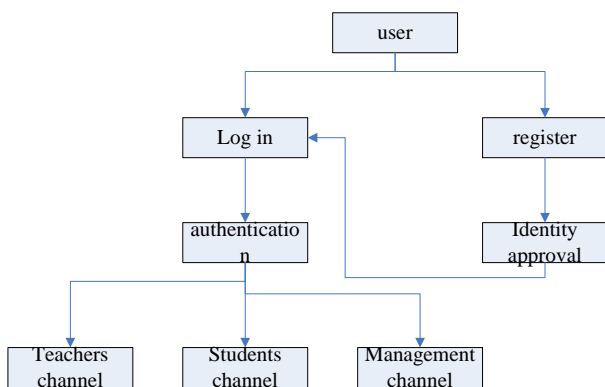
FIGURE 6 Landing module flow chart

Login authentication is the certification that all website users must go through. First, the user is required to fill in the user name and password on the page, for non-registered users, they need to register first. Then call database to search if there is a corresponding user. Code snippet is as follows:

```
mysql_conn();
Get Login Message ();
$today=date ("Y-m-d H:i:s");
$query="select id from user tbl /*
Where name=$name and password=$password
$result=my sql_query($query);
$numrows=mysql_num_rows($result);
If ($numrows==0)
Lonin Redirect (login Res ID)
Else {
$row=mysql_fetch_array($result);
$id=$row [0];
$query=" update usertbl set lastlogin=$today where id=$id";
$result=mysql_query($query);
```

```
Create ID Sesssion()
Lonin Redirect (login ResID)
}
```

4.3 ONLINE LEARNING MODULE DESIGN

When users make resources query according to their own needs, they first need to input the relevant CD information, and system will perform query information as requested. The main design code is shown below:

```
Protected void btn Search_Click(object sender, Event Args e)
{
String str =
Configuration Manager. Connection Strings ["guangpansystemConnectionString1"]. To String ();
Sql Connection con = new Sql Connection (str);
con.Open();
String book = this.txt Book name. Text. Trim ().To String ();
Sql Data Adapter sda = new Sql Data Adapter ("select * from guangpan where name like '%" + guangpan + "%'", con);
Data Set set = new Data Set ();
sda. Fill(set);
This.GridView1.Data Source = set;
This.GridView1.Data Bind ();
}
```

4.4 PERFORMANCE MANAGEMENT MODULE DESIGN

The performance needs to be managed mainly comes from test performance, task performance and course grade three aspects. Test performance will be automatically typed in by the system when students complete the test according to the test results. Task performance will be typed in by teachers when reviewing the students' work. Course grade is typed in by the teacher according to a comprehensive evaluation of test performance, homework completion as well as the usual classroom performance of the students. Therefore, written examination results of traditional teaching can also be input and managed in this module. Score type-in work is done by the administrator while the teachers and students only have the permission of query results, as shown in Figure 7. After the administrator logs in the performance management module, he can input the scores based on class information and course information. Also, operations such as modify and delete can be performed on the already type in grades and results output statements and reports can be produced.
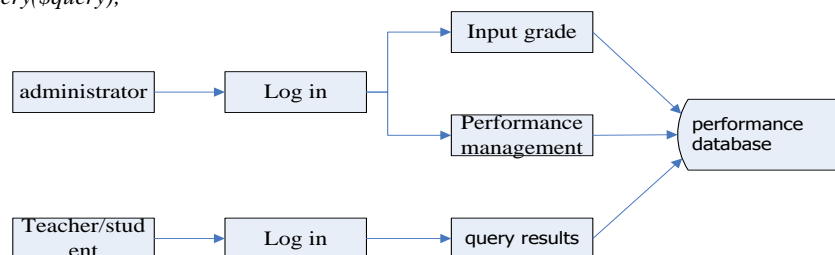
FIGURE 7 Performance management module structure chart

5 Conclusion

With the development of network technology and the popularization of network application, online teaching has become a focus of the current education. As people begin to call for lifelong education and the quality of life is improved. Higher requirements are put forward on the space, time and form of knowledge acquisition, which will bound to promote the further development of distance education. Modern distance education is a wonderful combination of traditional education and modern communication technology, and it will become the hot spot of education development in the 21st century.

In this paper, the advantages and disadvantages of distance education are discussed. A remote education software and hardware system based on B/S (Browser/Server) structure and the frame structure of relevant auxiliary teaching system are proposed and implemented. This paper applied the modular design method and divided the whole system platform into support system and supporting system, which plays a supplementary role in teaching according to different function of each module. In the design of support system in the remote education platform, it realized five main teaching functions of network classroom, live classroom, and courseware on demand, intelligent test and performance management. A real-time interactive teaching mode with an organic combination of live classroom and text interaction is put forward. For the support system, it realized remote education auxiliary functions of online homework, BBS message, resources download, etc., and online homework help is put forward which makes it possible for students to finish the homework online and teachers assign and mark homework online. This design realized the main function modules of the remote education information system.

Acknowledgment

References

[1] Jixin G, Huang D 2000 A new concept of teaching-online distance learning *China Distance Education* **01** 44-6 *(in Chinese)*
[2] Xingfu D 2001 The remote education *Beijing Peking University Press (in Chinese)*
[3] Danqing H 2000 National characteristics and development trend of open and distance education *China educational technology* **01** 57-9 *(in Chinese)*
[4] Qingyun H 2000 Research on development of distance education abroad *Shanghai Education Press (in Chinese)*
[5] Shengquan Y, Kekang H 1998 Teaching mode based on Internet. *China Educational Technology* **04** 58-61 *(in Chinese)*
[6] Abiteboul S, Benjelloun O, Manolescu I, Milo T, Weber R 2002 Active XML: Peer-to-Peer data and Web *Hong Kong Morgan*
[7] Ghandeharizadeh S, Sommers F, Joisher K, Alwagait E 2002 A document as a Webservice: Two complementary frameworks *XML-Based Data Management and Multimedia Engineering – EDBT 2002 Lecture Notes in Computer Science* **2490** 450-61
[8] Xiaoling G, Pengyu L 2004 Distance education systems *Beijing: China Aerospace Press (in Chinese)*
[9] Venna M 2003 XML [DB/DL].http://www-128.ibm.com/ developer works/cn/xml/x-seclay2/index.html.

Authors

Min Liu, Hebei Province, China

Current position, grades: lecturer.
University studies: Master degree in Software Engineering at Beijing University of Technology.
Scientific interest: computer technologies.

Jing Cao, Hebei Province, China

Current position, grades: lecturer.
University studies: Master degree in Engineering at Yanshan University.
Scientific interest: computer technologies.

Yanru Xue, Hebei Province, China

Current position, grades: lecturer.
University studies: Master degree at Ocean University of China in 2006.
Scientific interests: optical fiber communications.

Yinghua Yao, Hebei Province, China

Current position, grades: lecturer.
University studies: Master degree in Software Engineering at Beijing University of Technology in 2010.
Scientific interests: computer technologies.

Xuezuo Zhao, Hebei Province, China

Current position, grades: lecturer.
University studies: Bachelor degree was earned in Hebei Normal University in 2001.
Scientific interest: computer technologies.

Operation Research and Decision Making

A research on the intelligent multi-objective optimization problem based on wavelet theory and neural networks

Shunye Wang*

Department of Mathematic and Computational Science, Langfang Teachers University, Langfang 065000, China

Received 1 May 2014, www.cmnt.lv

Abstract

Aiming to solve the multi-objective optimization problem caused by wavelet multiresolution a1nalysis (MRA), the thesis improves the original multi-objective non-dominated genetic algorithm. After fast non-dominated sorting, the evolution of population is achieved through particle swarm optimization (PSO). In this way, the thesis realizes a more effective, organic combination of the multi-objective optimization problem and neural networks. MRA is a natural fit for the multi-objective optimization problem. The ability of neural networks to deal with complex errors is improved through error decomposition based on different wavelet decomposition scales.

Keywords: multi-objective optimization problem, wavelet theory, PSO, neural networks

1 Introduction

This paper explores the combination of neural network and wavelet theory, focusing on the processing of the noisy data and improves the existing wavelet multi-resolution analysis by combing the learning methods of neural network. The improvements are made on the multi-objective optimization problems caused by the multi-resolution analysis based on the superiority of the non-dominated particle swarm optimization.

Firstly, we turn the problem of predictive modelling into multi-objective optimization problem by using wavelet multi-resolution analysis. Then, we combine the non-dominated quick sorting and particle swarm optimization in dealing with the multi-objective optimization problem. The non-dominated multi-objective particle swarm optimization can converge quickly and it also keeps another advantage of particle swarm, that is, it only needs to know the objective function value instead of the derivative value, thus relaxing the requirements on the objective function. Therefore, the neural network using the parameters of particle swarm optimization network has distinguished advantages. In addition, it is quite easy to realize this kind of combination and it just needs to set the link weights and the output thresholds as the position vectors of the particle, turning the problem into the wavelet multi-resolution expression for the study errors. The network thresholds and weights iterate according to particle swarm optimization and search the optimal value. In this way, it not only avoids the neural network from making assignments randomly to the weights and thresholds in the initial stage, but it also avoids considering the error back propagation as a method to revise the rules so as to trap the network into the local minima values; therefore, it enhances the network stability. The

adjustments to the parameters of the wavelet neural network depends on the continuous search of the particle and indirectly determine the various link weights and thresholds of the neural network by updating the velocity and position vectors of the particle. Finally, the optimal value of the parameters of the wavelet neural network is determined by recording the optimal position of the particle.

2 Wavelet multiresolution analyses

Wavelets are used in a wide range of applications such as signal analysis, signal compression, finite element methods, differential equations, and integral equations. In the following we will discuss the limitations of traditional basis expansions and show why wavelets are in many cases more efficient representations. A mathematical treatment of second generation wavelets as well as an example will be provided.

1) Localization in space The Fourier transform is localized in frequency but the global support of the basic functions prevents localization in space. For many applications in particular the local behaviour of signals is of interest.

2) Faster transform algorithms. In recent years, the advance of data acquisition technology outpaced the available computing power significantly making the Fast Fourier Transform with its O (n log n) complexity a bottleneck in many applications.

3) More flexibility Traditional basis expansions provide no or almost no flexibility. It is therefore usually not possible to adapt a representation to the problem at hand. An important reason for this lack of flexibility is the orthogonal nature of traditional basis expansions.

---

* *Corresponding author* e-mail: Shunyew@163.com

Operation Research and Decision Making

4) Arbitrary domains Traditional basis representations can only represent functions defined of Euclidean spaces R^{n2}. Many real-world problems have embeddings $X \subset R^n$ as domain and it is desirable to have a representation which can be easily adapted for these spaces.

5) Weighted measures and irregularly sampled data traditional transforms can usually not be employed on spaces with weighted measures or when the input data is irregularly sampled.

These limitations motivated the development of wavelets. Many different fields such as applied mathematics, physics, signal processing, and computer science provided contributions and today both a thorough mathematical theory and fast and practical algorithms exist [1].

Wavelets can be categorized into discrete (DWT) and continuous (CWT) wavelet transforms. To speak in broad terms, the basic functions of DWTs are defined over a discrete space which becomes continuous only in the limit case, whereas the basic functions of CWTs are continuous but require discretization if they are to be used on a computer; see for example the book by Antoine et al. for a more detailed discussion of the differences. In signal compression applications mostly discrete wavelets are employed, whereas for signal analysis typically continuous wavelets are used [2].

2.1 CONTINUOUS WAVELET TRANSFORM

The Continuous Wavelet Transform (CWT) is a time-frequency representation of signals that graphically has a superficial similarity to the Wigner transform.

A wavelet transform is a convolution of a signal s(t) with a set of functions which are generated by translations and dilations of a main function. The main function is known as the mother wavelet and the translated or dilated functions are called wavelets. Mathematically, the CWT is given by:

$$W(a,b) = \frac{1}{\sqrt{a}} \int s(t) \psi \left(\frac{t-b}{a} \right) dt , \qquad (1)$$

where b is the time translation and a is the dilation of the wavelet.

From a computational point of view it is natural to use the FFT to compute the convolution which suggests that the results are dependent on proper sampling of $s(t)$.

When the mother wavelet is complex, the CWT is also a complex valued function. Otherwise the CWT is real. The squared magnitude of the CWT $|W(a,b)|^2$ is equivalent to the power spectrum so that a typical display (image) of the CWT is a representation of the power spectrum as a function of time offset b. One should note however that the precise form of the CWT depends on the choice of mother wavelet y and therefore the extent of the equivalency between the squared magnitude of the CWT and the power spectrum is application dependent.

The CWT operation is implemented using both the FFT and the discrete sum approach. You can use either one to get a representation of the effective wavelet using a delta function as an input [3].

2.2 DISCRETE WAVELET TRANSFORM

The Discrete Wavelet Transform (DWT) is similar to the Fourier transform in that it is a decomposition of a signal in terms of a basis set of functions. In Fourier transforms the basis set consists of sines and cosines and the expansion has a single parameter. In wavelet transform the expansion has two parameters and the functions (wavelets) are generated from a single "mother" wavelet using dilation and offsets corresponding to the two parameters.

$$f(t) = \sum_a \sum_b c_{ab} \psi_{ab}(t) , \qquad (2)$$

where the two-parameter expansion coefficients are given by:

$$c_{ab} = \int f(t) \psi_{ab}(t) dt \qquad (3)$$

and the wavelets obey the condition:

$$\psi_{ab}(t) = 2^{\frac{a}{2}} \psi(2^a t - b) . \qquad (4)$$

A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like one might see recorded by a seismograph or heart monitor. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. Wavelets can be combined, using a "reverse, shift, multiply and integrate" technique called convolution, with portions of a known signal to extract information from the unknown signal [4].

For example, a wavelet could be created to have a frequency of Middle C and a short duration of roughly a 32nd note. If this wavelet was to be convolved with a signal created from the recording of a song, then the resulting signal would be useful for determining when the Middle C note was being played in the song. Mathematically, the wavelet will correlate with the signal if the unknown signal contains information of similar frequency. This concept of correlation is at the core of many practical applications of wavelet theory.

As a mathematical tool, wavelets can be used to extract information from many different kinds of data, including - but certainly not limited to – audio signals and images. Sets of wavelets are generally needed to analyse data fully. A set of "complementary" wavelets will decompose data without gaps or overlap so that the decomposition process is mathematically reversible. Thus, sets of complementary wavelets are useful in wavelet based compression/decompression algorithms where it is desirable to recover the original information with minimal loss [5].

In formal terms, this representation is a wavelet series representation of a square-integrable function with respect to either a complete, orthonormal set of basis functions, or an over complete set or frame of a vector space, for the Hilbert space of square-integrable functions.

Wavelet theory is applicable to several subjects. All wavelet transforms may be considered forms of time-frequency representation for continuous-time (analogue) signals and so are related to harmonic analysis. Almost all practically useful discrete wavelet transforms use discrete-time filter banks. These filter banks are called the wavelet and scaling coefficients in wavelets nomenclature [6]. These filter banks may contain either finite impulse response (FIR) or infinite impulse response (IIR) filters. The wavelets forming a continuous wavelet transform (CWT) are subject to the uncertainty principle of Fourier analysis respective sampling theory: Given a signal with some event in it, one cannot assign simultaneously an exact time and frequency response scale to that event. The product of the uncertainties of time and frequency response scale has a lower bound. Thus, in the scale gram of a continuous wavelet transform of this signal, such an event marks an entire region in the time-scale plane, instead of just one point. Also, discrete wavelet bases may be considered in the context of other forms of the uncertainty principle.

Wavelet transforms are broadly divided into three classes: continuous, discrete and multiresolution-based.

As an important part of wavelet analysis, MRA is widely applied especially to image processing. It is known to all that MRA can be applied to neural networks. To put it simply, the object information described by MRA is decomposed into a series of infinite orthogonal spaces. Thus comes the problem - which spaces (or degrees) of object information do people want to acquire? Below we analyse the concept of MRA. Its expression is composed of closed subspaces $V_j, W_j \in L^2(R)$ that meet the following conditions [7]:

1) $V_j = \text{span}\{\phi_{j,k} \in Z\}, j \in (0, +\infty)$,

2) $V_j = \text{span}\{\psi_{j,k}, j, k \in Z\}$,

3) $V_{j+1} = V_j \oplus W_j$,

where x, z, R, and $f(\cdot)$ stands for one-dimensional system input, the set of integers, the set of all real numbers, and system model function, separately. V_j and W_j constitutes a mutually orthogonal space. The respective orthogonal basis within the two can be described as $\phi_{j+1}(x) = 2^{-j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{-j/2}\psi(2^j x - k)$. Meanwhile, the respective base of the mutually orthogonal space represents the wavelet scale and the wavelet function.

$f(x)$ can be expressed as follows if it meets the condition $f(x) \in L^2(R)$:

$$f(x) = \sum_{k \in Z} \phi_{j_{max},k}, f > \phi_{j_{max},k}(x) + \sum_{j_{max} \geq j \geq j_0, k \in Z} <\psi_{j,k}, k > \psi_{j,k}(x), \tag{5}$$

where j_{max} and j_0 means the upper limit and the lower limit of resolution; $< ..., ... >$ represents the inner product in the orthogonal space of the function. Theoretically, it is possible to realize infinite approximation to a function. As shown in Equation (5), a signal is composed of two parts—scaling function and wavelet function. We divide them into the approximation part and the detailed part [8, 9].

3 Non-nominated multi-objective particle swarm optimization

The multi-objective particle swarm optimization constructs non-dominated solution set based on Pareto dominance relation and preserves the non-dominated solution set currently found by using archive. To use ε dominance concept to update the archive can give the algorithm with good distributive; fasten the convergence of the solutions by improving the selection methods of global extremum and individual extremum of the algorithm and adopting new particle update strategies; add in self-adaptive mutation particle to avoid being trapped into local Pareto optimal solution and propose non-dominated set construction method based on quicksort to accelerate the operational efficiency of the algorithm[8].

3.1 THE MAIN PROCESSES OF MULTI-OBJECTIVE PARTICLE SWAM OPTIMIZATION

The main framework of the algorithm is as follows:
Step 1. Initialize the population.
Step 2. Make iteration loop till the Maxgen.
1) Update the velocity of every particle in the particle swarm according to Equation (6). In Equation (6), the constants C_1 and C_2 control the effect the individual extremum and the global extremum play on the particle update. Besides, the parameter W is named the inertia weight and R_1 and R_2 are the random numbers among [0,1].

$$Vel[i] = W \times Vel[i] + C_1 R_1 \left(P_{best}[i] - Pop[i]\right) + C_2 R_2 \times \left(G_{best}[i] - Pop[i]\right). \tag{6}$$

2) According to the update velocity of every particle obtained from the last step; recalculate the new variable value based on Equation (7) and replace the particle with the individual extremum if the new position is dominated by the individual extremum of this particle.

$$Pop[i] = Pop[i] + Vel[i]. \tag{7}$$

Step 3. Implement mutation operation on the particle swarm with self-adaptive mutation probability.
Step 4. Calculate the objective function value of every particle.

321

Step 5. Solve the non-dominated set *Npop*: find all the non-dominated particles in the non-dominated set through quicksort.

Step 6. Update the individual extremum. Here dominance concept is used. As for every particle in the swarm, if the current position of particle *i* dominates its individual extremum position, then update its individual extremum.

$$P_{best}[i] = Pop[i]. \tag{8}$$

Step 7. Update the archive Rpop, insert the non-dominated set of the swarm in the archive based on the ε-dominance relation and use ε-dominance concept in the update strategy.

Step 8. Update the global extremum and select one particle from the archive *Rpop* as the global extremum of particle *i*.

The archive *Rpop* preserves the optimal result of every generation of operation and after the algorithm is over, all the particles in *Rpop* is the final results of the algorithm [9].

3.2 THE KEY OPERATORS OF MULTI-OBJECTIVE PERTICLE SWARM OPTIMIZATION

3.2.1 Archive

The purpose to set an archive apart from the swarm is:

a) To preserve the non-strong ε- domination found in the iteration so as to lead the algorithm to get closed to the Pareto optimal region more quickly;

b) To exist as the candidate set of the global extremum of every particle in the swarm;

c) The solution set preserved by the archive is the result of the algorithm.

Therefore, in order to make the archive play the dominant role and the final solution set preserve excellent distributive, it should be considered how to make the externally-centralized particles distributed uniformly in the archive update. After some analysis, we adopt ε-dominance relation. ε-dominance is defined:

Make $f, g \in \mathbb{R}^{+m}$; name f ε-dominance $g(\varepsilon > 0)$; mark $f \succ_{\varepsilon} g$, if:

$$\forall i \in \{1,...m\} : \lfloor f_i / \varepsilon_i \rfloor \leq \lfloor g_i / \varepsilon_i \rfloor, \tag{9}$$

$$\forall i \in \{1,...m\} : \lfloor f_j / \varepsilon_j \rfloor \leq \lfloor g_j / \varepsilon_j \rfloor. \tag{10}$$

Figure 1 demonstrates the concepts of dominance relation and ε-dominance relation. It can be seen that in ε-dominance relation, the solution space is divided into w meshes and the dominance domain of the individual f expands than the normal dominance domain. Use it in the multi-objective particle swarm optimization to make the solution set preserve better distribution.
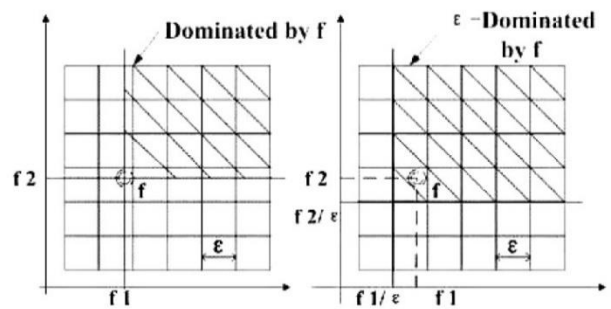
FIGURE 1 Dominance and ε-Dominance

3.2.2 Individual update

When the particle uses Equations (6) and (7) recalculates its decision variable, the particle actually has found a new solution and a new position in the objective function space.

In the experiment, we find the test problems in the local Pareto optimal domain. During the algorithm operation, only a few particles of every generation can jump out from the local Pareto optimal domain and they are still trapped into the local solutions as a whole. More cycle indexes are needed to converge to the global Pareto optimal domain or it may not converge at all. In order to solve this problem and accelerate the convergence, we will compare its new position with the position of the individual extremum when we update the individual position:

a) If the new position is dominated by the individual extremum of this particle, then replace the particle with its individual extremum;

b) Otherwise, it remains unchanged in the new position.

The above comparison and revise make the entire swarm jump out of the local Pareto optimal domain in time and ensure that the whole swarm converges quickly to the global Pareto optimal domain [10].

3.2.3 The Selection of the individual extremum and the global extremum

The selection of the individual extremum and global extremum are greatly important in the particle swarm optimization and it will directly affects the convergence rate of the algorithm and the distributivity of the solution set for the multi-objective optimization problem.

As for the individual extremum, the selection of the individual extremum is different from the single-objective optimization in the multi-objective optimization problem and we will judge the optimal position a particle finds based on the dominance relation.

In terms of the global extremum, the archieve *Rpop* is our candidate set. When starting the algorithm, set the global extremum of every particle as itself and conduct global extremum update on every generation. Randomly select a particle in the archive for every particle *i* in the swarm. If this candidate particle dominates the global extremum of the particle *i*, then set the candidate particle as the new global extremum of the particle *i*; otherwise, we

implement global extremum mutation on particle. As a matter of fact, we have proposed the extremum mutation method, namely set an extremum mutation probability. Produce a random number among [0, 1] for particle *i*. If the value of the random variable is smaller than the extremum mutation probability, then randomly select a particle from the archive to replace the current global extremum of the particle *i*.

3.2.4 The Design of the Disturbance Operators

One characteristic of particle swarm optimization is that it has fast convergence; however, it may also lead to being trapped in the local optimal solution; therefore, disturbance is needed to be implemented on the particles to make them jump out of the local optimal domain.

We have designed something similar to the mutation operators in the genetic algorithm to solve this problem. In the early operation of the algorithm, search the entire objective space. With the continuous evolution of the particles in the swarm, it needs to reduce the number of particles participating in the mutation. The mutation probability *Pmutation* is calculated according to Equation (11) and *Currentgen* is the current operational algebra.

$$Pmutation = 1 - Currentgen / Maxgen .\qquad(11)$$

For every particle in the swarm, the variable *m_random* is given a random number among [0,1]. If *m_random* is smaller than the mutation probability *Pmutation*, then conduct non-uniform mutation to that particle, namely to make mutation randomly on one of the n-dimensional decision vectors according to Equation (12). The value of θ represents that the forward direction of the mutated particle is the same as or opposite to its original direction; μ stands for the scope of velocity change and the random value of θ is ±1. In order to jump out from the local optimal domain through mutation, the velocity of the particle must be accelerated and μ should be bigger than 1. In the experiment, we set: *μ*=3.

$$Pop[i] = \theta \times \mu \times (1 - m\_random)Vel[i] + Pop[i] .\qquad(12)$$

If the mutation operation makes the decision variable exceed the definition domain, then set that decision variable in the boundary of its definition domain.

3.2.5 Construct non-dominated set through QuickSort

The construction of non-dominated set is a significant step in the multi-objective evolutionary algorithm. In order the increase the efficiency to construct non-dominated set, we have proposed to construct non-dominated set algorithm based on quicksort. Select an individual *x* (we usually select the first individual) from the swarm in every cycle of the algorithm and compare the other individuals with *x*. Divide the swarm into two parts through one comparison: the latter part is the individuals dominated by *x* while the

former part is the individuals dominating *x* or irrelevant to *x*. If *x* is not dominated by any individual, then integrate *x* into the non-dominated set and repeat the above process to the former part until it is empty.

Figure 2 displays the position changes of the particles in the swarm after a comparison. The particles before 1 is the particles not dominated by it while those after it are dominated by it obviously, only the front particles can enter into the non-dominated set. In this way, it narrows down the comparison range and saves expense.
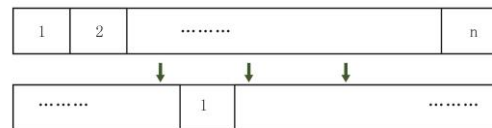
The particle not to be dominated by particle 1   The particle dominated by particle 1
FIGURE 2 Construct Non-dominated Set Based on QuickSort

4 Learning algorithm combing wavelet and neural network

4.1 WAVELET NEURAL NETWORK OF BP LEARNING ALGORITHM

The hidden layer neuron of wavelet neural network uses wavelet-based function as its network's activation function, which introduces scaling factor and translation parameter to the network in order to substitute original connection weight and output threshold, i.e. connection weight replaced by translation factor and output threshold replaced by scaling factor. Based on BP algorithm i.e. error back propagation algorithm, the training data in network is inputted to the neural network from input layer and then reach output layer via hidden layer, which is also called "pattern forward propagation"; the output value of neural network is compared with the ideal output and then the weights are adjusted starting from the output layer based on the thought of error reduction, which is called "error back propagation". The error accumulation after layer-by-layer transmission are just the signals that we want to catch at last and according to which each weight and threshold are adjusted in layer-by-layer inversed sequence based on error reduction[11]. The following example is a three-layer wavelet neural network (i.e. input, and output hidden layers), in which gradient descent method is used to find out errors and weights in order to get the variation of network weights and then the network is optimized by minimum mean square error objective function to adjust network weights and wavelet coefficient adaptively. The model of wavelet neural network is expressed as:

$$y_i^p(t) = f\left[\sum_{j=1}^{N} w_{ij}(t)\Psi_{a_j,b_j}\left(\sum_{k=1}^{m} w_{jk}(t)x_k^p(t)\right)\right] net_j^p =$$

$$\sum_{k=1}^{m} w_{jk}(t)x_k^p(t),\qquad(13)$$

$$\Psi_{a_j,b_j} net_j^p(t) = \psi\left[\frac{net_j^p(t) - b_j(t)}{a_j(t)}\right], \quad (14)$$

$$E(t) = \frac{1}{2}\sum_{p=1}^{M}\sum_{i=1}^{n}\left[d_i^p - y_i^p(t)\right]^2, \quad (15)$$

where t is iterative steps, x_k^p is the k-th input value of the p-th sample of input layer, $y_i^p(t)$ is the i-th output value of the p-th sample of output layer, d_i^p is the i-th desired output of the p-th sample, $\psi(x)$ is basic wavelet function, $E(t)$ is objective function and $f(x)$ is the purlin of output layer. In this paper, it is supposed that $f(x)=x$, $w_{ij}(t)$ is the

weight connecting the node j of hidden layer and the node i of output layer, $w_{jk}(t)$ is the weight connecting the node k of input layer and the node j of hidden layer, $a_j(t)$ and $b_j(t)$ are the scaling and translation coefficients of the node j-th of hidden layer respectively, m is the number of nodes of input layer, N is the number of the nodes of hidden layer, M is the number of input samples and n is the number of the nodes of output layer. In this way, after the connection weight w, threshold b and the scaling factor and translation parameter a, b of wavelet network are initialized and the input information is propagated forward, the actual ouput of the wavelet neural network can be calculated using current network parameters and the network can be optimized as follows:

$$\frac{\partial E(t)}{\partial w_{jk}(t)} = \frac{-\sum_{p=1}^{M}\sum_{i=1}^{n}\left(d_i^p - y_i^p(t)\right)w_{ij}(t)\Psi'_{a_j,b_j} net_j^p(t) x_k^p(t)}{a_j(t)}, \quad (16)$$

$$\frac{\partial E(t)}{\partial w_{ij}(t)} = -\sum_{p=1}^{M}\left(d_i^p - y_i^p(t)\right)\Psi_{a_j,b_j}\left(net_j^p(t)\right), \quad (17)$$

$$\frac{\partial E(t)}{\partial a_j(t)} = \frac{-\sum_{p=1}^{M}\sum_{i=1}^{n}\left(d_i^p - y_i^p(t)\right)w_{ij}(t)\Psi'_{a_j,b_j}\left(net_j^p(t)\right)\left(\frac{net_j^p(t) - b_j(t)}{a_j(t)}\right)}{a_j(t)}, \quad (18)$$

$$\frac{\partial E(t)}{\partial w_{ij}(t)} = \frac{-\sum_{p=1}^{M}\sum_{i=1}^{n}\left(d_i^p - y_i^p(t)\right)w_{ij}(t)\Psi'_{a_j,b_j}\left(net_j^p(t)\right)}{a_j(t)}, \quad (19)$$

Parameters adjustment equations are the following:

$$w_{jk}(t+1) = w_{jk}(t) - \eta\frac{\partial E(t)}{\partial w_{jk}(t)} + \mu\Delta w_{jk}(t), \quad (20)$$

$$w_{ij}(t+1) = w_{ij}(t) - \eta\frac{\partial E(t)}{\partial w_{ij}(t)} + \mu\Delta w_{ij}(t), \quad (21)$$

$$a_j(t+1) = a_j(t) - \eta\frac{\partial E(t)}{\partial a_j(t)} + \mu\Delta a_j(t), \quad (22)$$

$$b_j(t+1) = b_j(t) - \eta\frac{\partial E(t)}{\partial b_j(t)} + \mu\Delta b_j(t), \quad (23)$$

where, μ is a momentum coefficient and also the learning rate of wavelet neural network. Because of the use of derivative, BP algorithm is very dependent on objective function, which is to say, in case of different objective functions, the training process of BP algorithm is also unavailable even if the network structure is the same [12].

4.2 WAVELET NEUTRAL NETWORK BASED ON MULTI-SUBSWARM LEARNING ALGORITHM

4.2.1 Multi-subswarm particle swarm optimization algorithm

The main steps based on multi-subswarm particle swarm optimization algorithm include the following:
Step 1. Set up parameter set and assign values to particle swarm randomly.
Step 2. Calculate the fitness values of all current particle S and divide them into two subswarms according to the fitness value.
Step 3. The two subswarms evolve according to their own strategies:
a) Each particle of the subswarm S_2 breeds some optional off springs according to the pre-set growth rate ρ_2, the fitness function values of those optional offsprings are calculated, and the top L particles from higher fitness value to lower fitness value evolve further and other optional particles are not selected;
b) Each particle of the subswarm S_1 breeds an offspring, the H off springs are selected for further evolution and their fitness function values are calculated.

Step 4. The results of the evolution of the two subswarms are combined and the optimal value is updated.

Step 5. Whether to meet END condition is checked. If one of the following is met, i.e the biggest preset evolution algebra and the requirement for accuracy, calculation ends; otherwise skip to Step 2.

It is supported that $S = \{X_i, i = 1, 2, L\ N\}$ represents the subswarm of N particles and X_i is the i-th particle of the subswarm, so the fitness values of all the particles are arranged in ascending order as follows:

$$F = \left\{ \left(f_1, f_2, L\ f_N\right) \mid f_1 \leq L \leq f_k \leq L\ f_N \right\}. \tag{24}$$

$S_1 \cup S_2 = S, S_1 = \Phi$. $H + L = N; S_1, S_2$ are the subswarms of H and L particles respectively where: $H = N - \left[L_{\max} \times rand()\right]$. $L_{\max}$ represents the biggest number of particles in S_2 and $rand()$ is the random number of $[0,1]$. $p_2 = \dfrac{l_{\max}}{N}$ result shows that the value range is 0.02~0.10 to the best.

In consideration of the number of the offsprings of subswarms, this paper introduces the growth rate ρ_1, so the scale H_1 of the offsprings of subswarm S_1 can be expressed as: $H_1 = \rho_1 \times H$

Quantities of experiments show that it is proper to set the growth rate ρ_1 of subswarm S_1 to be 1, so there is: $H_1 = \rho_1 \times H \rho_1 = 1$.

The evolution of subswarm S_1 is subject to the evolution process of basic particle swarm algorithm, i.e. each particle only has one offspring and all the off springs are reserved as the parents of the next generation. The evolution of subswarm S_2 is comparatively complicated, mainly because local optimum is avoided based on this subswarm and each particle needs to produce several off springs. In convenience for description, H_2 is the scale of the off springs of subswarm S_2 and ρ_2 is the growth rate of subswarm S_2, so there is: $H_2 = \rho_2 \times L, \rho_2 \geq 1$.

4.2.2 Wavelet neural network based on multi-subswarm learning algorithm

The number of hidden layer neurons is closely related to the approaching ability of network, that is, the more number of neurons, the better of the approaching property of the network. However, in reality, other factors need to be considered like the complexity, so it does not mean the more the better. Therefore, in terms of wavelet neural network, if it can adjust the number of its own neurons dynamically and adaptively, it will increase the scope of application of wavelet neural network. As the learning process of wavelet neural network can be treated as the adjustment process of hidden layer neuron parameter, the position vector of each particle can be defined as:

$$present(i) = \left[w_i, a_i, b_i, \lambda_i\right], i = 1, 2, L, N. \tag{25}$$

The model of optimizing wavelet neural network with multi-subswarm particle swarm is the following:

1) Determine particle dimensions.

In case of one neuron, 4 variable parameters can be used to determine one network output and all the parameters that can be used to determine one ouput are defined as one particle. In this way, if there are N hidden layer neurons, the network will need 4N parameters. As the number of parameters corresponds to the dimensions of searching space, particle dimensions are 4N dimensions.

2) Initialize particle subswarm.

Initialize particle subswarm with random value assignment before network training.

3) Calculate the output of neural network.

Standardized processing is made to the training data sample of wavelet neural network and Mexican Hat Function is used as the activation function of hidden layer neurons. The position and velocity vectors are updated with multi-subswarm particle swarm optimization algorithm and then the actual ouput value and error function value of network can be obtained according to particle position x and training sample.

4) Judge generation quantity.

If the upper limit of generation quantity is not reached, particle swarm shall be updated, and step (3) is taken, or further step can be taken;

5) Judge whether to stop training.

If the optimal particle cannot meet desired error threshold, further step can be taken, or the process shall end;

6) Judge whether to adjust structure.

If the particles near the upper limit of generation quantity can't meet desired variation, i.e. the particles only vary with a small range, the number of neurons shall be increased and Step 2 shall be taken; or, without adjustment of network structure, the particle swarm can be updated and Step 3 can be taken;

The training process of multi-subswarm structure particle swarm wavelet neural network is as follows:
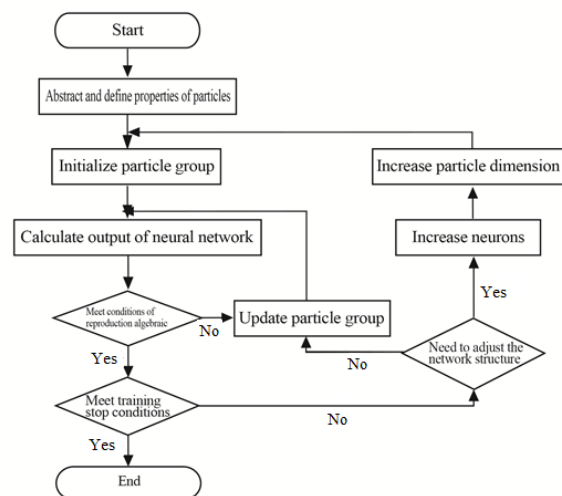
FIGURE 3 The Training Process of Multi-subswarm Particle Swarm-optimized Wavelet Neural Network

5 Simulation examples

The thesis chooses Mackey-Glass and the sunspot chaotic time series as the simulation objects. Normalization processing is still an essential step before the algorithm begins. The processing standards should be conducive to the effective implementation of algorithms. Normalization is also closely related to the allocation of wavelet nodes. In this simulation, the wavelet nodes are strictly limited to the domain of definition $[0,1]^n \, (n < 4)$.

5.1 LOGISTIC SYSTEM

Under the condition of $0 \le t \le \tau$, make $x(t) = 0.9$. The integration step is 0.01, the time delay τ is 30, and the sampling time interval is 0.06. There are 1,000 data points to be tested and trained. We can select from the generated data at random. Add Gaussian noise of 10% and 20% respectively to the original samples to yield noise data.

The Equation of the Logistic system is:

$$x_{i+1} = 4x_i (1 - x_i).$$ (26)

We continue the one-step prediction and divide the 1,000 sample points into two parts. The first half is used to train the network, and the second half is used to test the network. According to literature, expand the wavelet network for simulation based on ANOVA. Set the embedded dimension of phase space at 6, where $x_{l_i} = y(t - l_i), i = 1, 2, 3, 1 \le l_1 \le l_2 \le l_3 \le 6$. The note function of the network is still one-, two-, or three-dimensional. Besides, its number and exponents decrease progressively $(g_{l_1}^{[1]}(\cdot), g_{l_1 l_2}^{[2]}(\cdot), g_{l_1 l_2 l_3}^{[3]}(\cdot))$. Equations (28) and (29) respectively stand for the one-dimensional form and the multi-dimensional form of the wavelet.

The network model is as follows:

$$y(t) = g(x_1, x_2, x_3, x_4, x_5, x_6) =$$
$$\sum_{1 \le l_1 \le 6} g_{l_1}^{[1]}(x_{l_1}) + \sum_{1 \le l_1 \le l_2 \le 6} g_{l_1 l_2}^{[2]}(x_{l_1}, x_{l_2}) = ,$$ (27)
$$\sum_{1 \le l_1 \le l_2 \le l_3 \le 6} g_{l_1 l_2 l_3}^{[3]}(x_{l_1}, x_{l_2}, x_{l_3})$$

$$\psi^{[1]}(x) = (1 - x^2)e^{-\frac{1}{2}x^2},$$ (28)

x is one-dimensional input.

$$\psi^{[n]}(x) = (n - \|x\|)e^{-\frac{1}{2}\|x\|^2},$$ (29)

x is multi-dimensional input.

We train the network by combining fast non-dominated sorting and PSO. The optimization objective is described as:

$$\|e_1(x)\| = \left\| \sum_{k \in Z} \phi_{1,k}, e_0 > \phi_{1,k}(x) \right\|.$$ (30)

Based on the above objective function form, we can obtain several objective functions by changing the scale of wavelet resolution. Here we adopt two scales starting from the original scale and get two objective functions. Below are the experiment results.

(a)Original data

(b)Add 10% of the noise
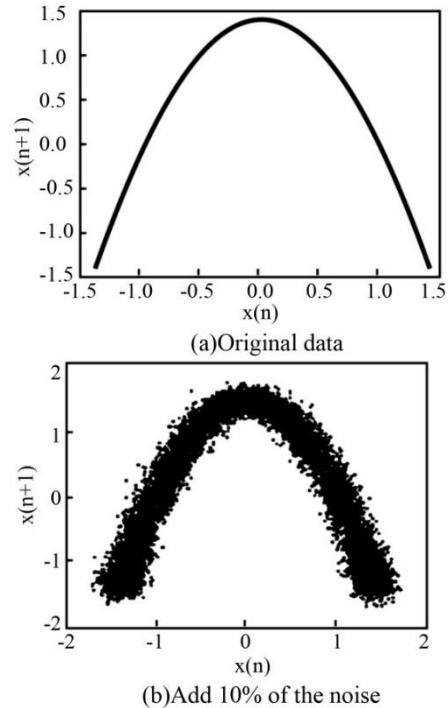
FIGURE 4 Two-dimensional reconstructed attractor of Logistic map
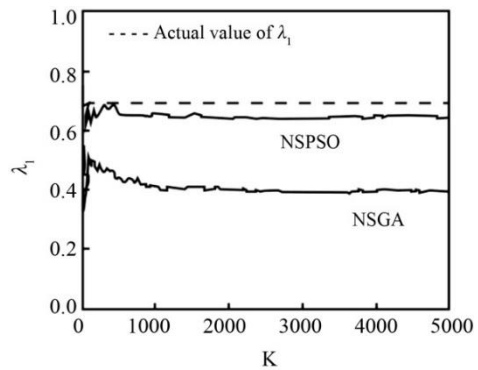
FIGURE 5 The convergence of Lyapunov spectrum of Logistic map time series with 10% noise

Figures 4a and 4b respectively represent the two-dimensional reconstructed attractor before the addition of nosie and after the addition of 10% noise. We can know that Logistic mapping is a parabola in two-dimensional phase space, but the results become a "parabolic band" when it contains noise.

From the simulation results of NSGA, and NSPSO, as the noise increases, the advantage of multi-resolution algorithm becomes more obvious. This is mainly because the multi-objective optimization of multiresolution learning algorithm can cover noise of more frequencies. More importantly, it is notable that the resolution of PSO is better than that of the genetic algorithm.

Operation Research and Decision Making

5.2 HÉNON SYSTEM

Secondly, calculate the time series generated by Hénon map, Figures 4a and 4b respectively represent before the addition of noise and after the addition of 15% noise. As shown, we find that the latter one made the reconstructed attractor lose the detailed traits completely. Besides, we take the measures of NSGA as well as NSPSO so as to calculate the time series of the particles containing noise. Under the condition of $\Delta t = 1$, $N_s = 5 \times 105$, $\tau = 1$, $d = 2$, $\lambda_1 = 0.4185$, $\lambda_2 = -1.6225$, add Gaussian noise of 15% to the original samples to yield noise data.

The equation of Logistic system is:

$$\begin{aligned} x_{i+1} &= 4x_i\left(1-x_i\right) \\ y_{i+1} &= 0.3x_i \end{aligned}, \tag{31}$$

The wavelet function in the simulation example is Gauss wavelet. The optimization algorithms are the combination of fast non-dominated sorting and heredity as well as particle swarm. The network model and the fitness function expression of non-dominated PSO are as follows:

$$y(t) = g(x) = g\left(y(t-1), y(t-2),..., y(t-9)\right), \tag{32}$$

$$\|e_0\| = \|y(t) - g(x)\|, \tag{33}$$

$$e_1(x) = \sum_{k \in Z}\langle \phi_{1,k}, e_0 \rangle \phi_{1,k}(x) \tag{34}$$

(a)Original data

(b)Add 15% of the noise
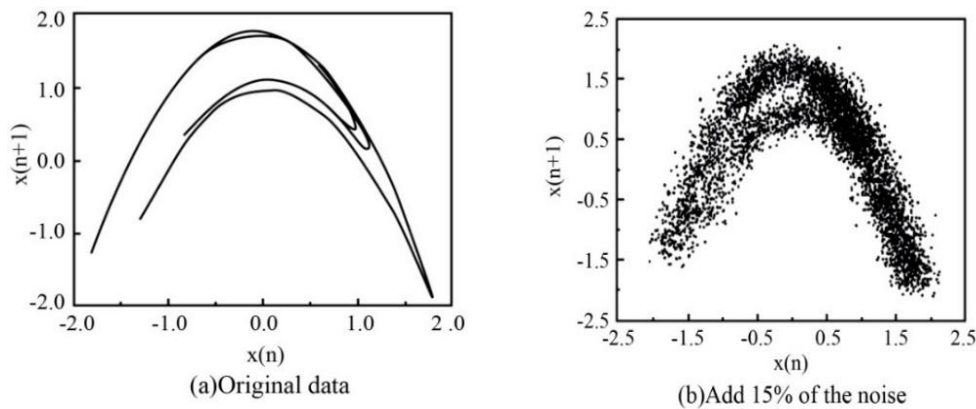
FIGURE 6 Reconstructed attractor of Hénon map

(a) Convergence graph of λ_1

(b) Convergence graph of λ_2
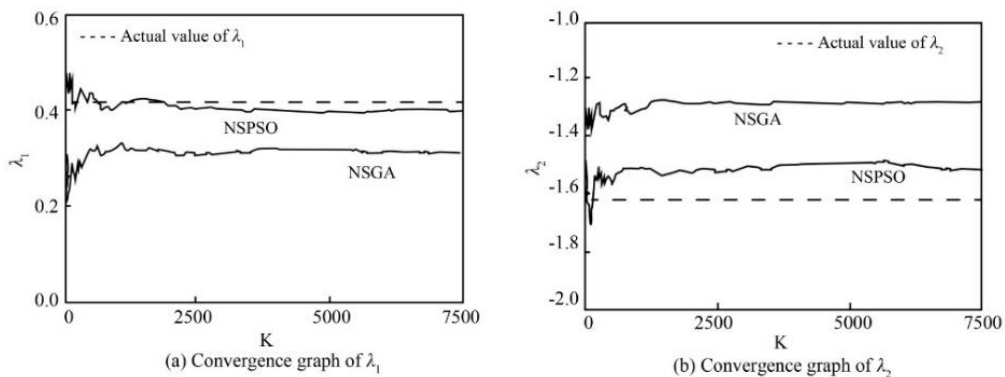
FIGURE 7 The convergence of Lyapunov spectrum of Hénon map time series with 15% noise

Figure 7 indicates that the introduction of MRA greatly improves the ability of wavelet network learning and noisy data prediction. It obviously proves the louder the noise, the higher the efficiency of MRA algorithms. Thus, MRA has more important practical applications.

6 Conclusions

The thesis makes a research on noisy data processing through the combination of neural networks, particle swarm optimization and wavelet theory so as to improve the multi-objective optimization problem caused by

multiresolution analysis. The improvement is based on the superiority and popularity of PSO over genetic algorithms. Firstly, the thesis converts the predictive modelling problem to the multi-objective optimization problem by dint of MRA. After that, it adopts the training method integrating fast non-dominated sorting and PSO to deal with the multi-objective optimization problem. The simulation results of the experiment shows, the algorithm proposed in the thesis has an advantage when handling the problems of multi-objective optimization and errors.

327

References

[1] Al-Geelani N A, Piah M A M, Adzis Z, Algeelani M A 2013 Hybrid regrouping PSO based wavelet neural networks for characterization of acoustic signals due to surface discharges on H.V. glass insulators *Applied Soft Computing* **13**(12) 4622-32

[2] Liu H, Tian H, Chen C, Li Y 2013 An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization *International Journal of Electrical Power & Energy Systems* **52** 161-173

[3] Liu Z, Cao H, Chen Z, He Z, Shen Z Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings *Neurocomputing* **99** 399-410.

[4] Chatterjee A, Ghoshal S P, MukherjeeV 2011 Craziness-based PSO with wavelet mutation for transient performance augmentation of thermal system connected to grid *Expert Systems with Applications* **38**(6) 7784-94

[5] Tsai H, Jhuang Y, Lai Y 2012 An SVD-based image watermarking in wavelet domain using SVR and PSO Original Research Article *Applied Soft Computing* **12**(8) 2442-53

[6] Shrivastava N A, Panigrahi B K 2014 A hybrid wavelet-ELM based short term price forecasting for electricity markets *International Journal of Electrical Power & Energy Systems* **55** 41-50

[7] Daamouche A, Hamami L, Alajlan N, Melgani F 2012 A wavelet optimization approach for ECG signal classification *Biomedical Signal Processing and Control* **7**(4) 342-9.

[8] Kusetogullari H, Leeson M S, Kole B, Hines E L 2014 Meta-heuristic algorithms for optimized network flow wavelet-based image coding. *Applied Soft Computing* **14** Part C 536-53

[9] Guo J, Ding L, Luo H, Zhou C, Ma L 2014 Wavelet prediction method for ground deformation induced by tunneling *Tunnelling and Underground Space Technology* **41** 137-151

[10] Katunin A, Przystałka P 2014 Damage assessment in composite plates using fractional wavelet transform of modal shapes with optimized selection of spatial wavelets *Engineering Applications of Artificial Intelligence* **30** 73-85

[11] Subasi A 2013 Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders *Computers in Biology and Medicine* **43**(5) 576-86.

[12] Meng Y, Zou J, Gan X, Zhao L 2012 Research on WNN aerodynamic modeling from flight data based on improved PSO algorithm *Neurocomputing* **83** 212-21

Author

Shunye Wang, born in June, 1981, Hebei, China

Current position, grades: lecturer Department of Mathematics and Computational Science, Langfang Teachers University.
University studies: M.S degree in Computer Science and Technology, Computer Applications from the China University of Petroleum, China, 2003, 2006.
Scientific interest: machine learning and data mining.

Design and algorithm of supply chain network model based on uncertain environment

Yulan Zheng∗

School of Economics & Management, Fuzhou University, Fuzhou, Fujian, 350116, China

Abstract

Under the conditions of modern market economy and the increasing pressure of competition, a growing number of product life cycles are getting shorter, customer demand is constantly changing, this force the companies began to pay attention to their own internal management and supply chain management. Supply chain refers to the suppliers, manufacturers, distributors and end-users, and the network, which design both directions of material flow, information flow and capital flow [1]. In order to manage and use it better, People put network information technology in supply chain management, developed or purchased ERP implementation systems software, thus can share information between departments and enterprises. But problems have gradually exposed, faster access and processing does not produce a good decision automatically. So this paper under the premise of uncertainty, started studying supply chain's network model, establishing optimization model, design algorithms, therefore reducing the total costs, to avoid risks.

Keywords: uncertain environment, supply chain, model, design

1 Introduction

Supply chain management problem at first is the supply chain network design, to determine the configuration of supply chain and business processes that every level need to implement, is the primary concern, typical decision are as follows: supplier's selection, facilities' location, facilities' capacity, demand and supply distribution. Decision problem can be planned in two parts: Supplier's selection, facilities issues. If one of the issues optimized, it can have great benefits to the whole enterprise as a whole. If merchant, manufacturer and supplier in the supply chain can work together, manage and optimize the flow processes, inventory can be reduced, flexibility can increasing, so as to achieve the goal of low cost and high efficiency.

2 Effect of uncertainty environment on supply chain management

The source of uncertainty in supply and production processes is the initial supplies. Because material supply effects began to gradually pass to downstream members, and then expand the production assembly process and delivery time of the product, thereby indirectly affecting the customer's satisfaction to the product. Uncertainty of customer demand will step by step pass to upstream members along the supply chain, thus affecting the item's inventory and storage time, thus increased inventory costs. Sometimes customers demand uncertainty will cause serious impact on the enterprise, especially serious for

some products, which difficult to keep and remodel frequent enterprise.

With the development of networks, production and market information's storage, analysis and delivered can quickly process. Supply chain information management is no longer limited by time or space, so that information can be shared directly between suppliers, manufacturers, distributors, customers, it greatly reduces bias caused by the layer-by-layer transmission of information zoom, Therefore, information possess a certain amount of certainty. Supply chain management is not only require information sharing, the supply of materials is necessary, if there is no matching system of logistics management, supply chain does not have an operating base, while not able to achieve e-commerce "business function".

So in a sense, supply chain management, aimed at preventing the identification of possible adverse consequences for supply chain performance, ensure that the uncertainty of the process can be reduced to a minimum to let material flow, increase flexibility, and reduce the cost of purchasing and supply and trading, through the optimization of the supply chain to improve the competitiveness of enterprises.

3 Design of supply chain network

Supply chain network design is strategic supply chain management issues. In the supply chain network, we can see all abstract as a directed graph consists of nodes and arcs. We are setting up a model that divided nodes into four categories: suppliers, manufacturers, distributors, and

∗*Corresponding author* e-mail:zylfzdx@163.com

Operation Research and Decision Making

customers demand points, arc is represented by the links between nodes, to represent business processes.

Supply chain network design's purpose is to identify products from raw materials to customer requirements points are complete channel architecture. Supply chain network design in this article need to solve the primary problem that considering the uncertainty of design parameters, to determine supply chain structure when meet constraints, and supply chain performance in parameter perturbation can remain stable enough, so as to avoid market risks.

Major constraints are:

1. supplier for all factory shipments cannot exceed the supplier's capacity;

2. factory shipments for all manufacturers cannot exceed factory capacity;

3. vendor shipments for customer demands cannot exceed manufacturers' capacity;

4. the total amount of goods from suppliers not less than the factory needs;

5. obtained from the factory cannot be less than the amount of cargo sellers demand;

6. obtained from the distribution centre cannot be less than the amount of goods to customer needs [1].

3.1 ESTABLISH MATHEMATICAL MODEL

Supply chain network design is a long-term strategic decision, because too few valid historical data can be used, it is difficult for policy makers using historical data to draw

the exact value of the parameter, and it is difficult to determine the parameters of the probability distribution function, you can use scenario analysis to examine the uncertainties "to reproduce."

First of all, the basic premise of establishing mathematical model is:

1. Indirect costs incurred by the supplier product can be obtained by using the activity-based costing method:

$$\varpi = \sum_t \sum_f \tau_t u_f x_{tf}$$, t is vendor performance evaluation

index number; f is cost driver number; u_f is the unit cost of cost drivers; τ_t is number of supplier performance index; x_{tf} is number of cost driver caused by supplier performance index.

2. Every customer area can only get delivery services from one seller. Through these premises, we can completely define the following method:

a) The main sources of uncertainty--customer demand and costs;

b) Parameter uncertainty--different value ranges;

c) Combine the value ranges of different situations, judge the rationality and form various scenarios;

d) Equal probability distribution of each scenario.

Based on the principle of "first divide later combine" we respectively established supplier selection and decision-making model (Table 1), after consolidation we set up supply chain network design model based on robust optimization theories.

TABLE 1 Supplier selection and decision-making model

	Collection		Index number		Design variable		Controlled variable
I	Supplier collection	i	SupplierNumber, i∈I	X_{im}	Whether choose supplier i provide raw material m; Yes: $X_{im} = 1$; No: $X_{im} = 0$	G^s_{ijm}	Number of raw material m that supplier i purchased from manufacturer j under the s scenery
J	Manufacturer collection	j	Manufacturer number, j∈J	Y_j	Whether establish manufacturer; Yes: $Y_j = 1$; No: $Y_j = 0$	Q^s_{jpk}	Volume of product p from manufacturer j to marketing centre k under s scenery
K	Marketing centre collection	k	Marketing centre number, k∈K	Z_k	Whether set up marketing centre k; Yes: $Z_k = 1$; No: $Z_k = 0$	R^s_{knp}	Volume of product p from marketing centre k to customer n under s scenery
N	Customer area collection	n	Customer area number, n∈N	XY_{kn}	Whether marketing centre k provide product for customer area n; Yes: $XY_{kn} = 1$ No: $XY_{kn} = 0$		
M	Raw material collection	m	Raw material number, m∈M	FC_j	Capacity of manufacturer j		
P	Product collection	p	Product number, p∈P	DC_k	Capacity of marketing centre		
S	Possible scenarios collection	s	Scenarios number, s∈S				

3.2 INTEGRATE SUPPLY CHAIN NETWORK DESIGN MODEL

After establishing a sub-model, we began to integrate supply chain network design model, before establishing model we have to defined first: $\forall s \in S$, ξ_s is the supplier's chosen raw material costs and facilities decision cost'

summation, then: $\xi_s = \xi^s_{VS} + \xi^s_{FL}$. For each specific scenario s, the parameter of the model is determined, here supply chain network design problems are common problems of the deterministic optimization problem, for every deterministic optimization problems, their numerical are differences, the rest of the structure is the same,

$\forall s \in S$, the objective function value is ξ_s^* after optimizing.

Make the parameter $\omega > 0$, if variable $x = \left\{ X_{im}, Y_j Z_k XY_{kn} FC_j DC_k \right\}$ is the feasible solution to all the deterministic optimization problems ND_s, then the objective function value to ND_s is $\xi_s(x)$.

Only if $\forall s \in S$, $\dfrac{\xi_s(x) - \xi_s^*}{\xi_s^*} \le \omega$, x is robust solution to supply chain network design problem.

$\forall s \in S$, left in the equation above is known as the relatively regretted, absolutely regretted is determined by $\xi_s(x) - \xi_s^*$, relatively regretted and absolutely regretted can change with each other by multiply or division the constant ξ_s^* .

When the probability of each scenario is different, you can define different regret limited coefficients ω.

From the above definition, we can know there can be more robust solutions of supply chain network design, and robust optimization is aimed at finding best robust solutions. Therefore, we can establish integrate robust optimization model (ROM) for supply chain network design ρ_s − probability of S scenery, ω − regret limited coefficients.

$$
\begin{aligned}
Min\xi = \sum_s \rho_s \xi_s , \\
\xi_s = \sum_i \sum_m \varpi_{im}^s X_{im} + \sum_j v_j^s Y_j + \sum_k w_k^s Z_k + \\
\sum_i \sum_j \sum_m (\alpha_{im}^s + \mu_{ijm}^s) G_{ijm}^s + \sum_j \sum_k \sum_p (\beta_{jp}^s + \eta_{jkp}^s) Q_{jkp}^s + \\
\sum_k \sum_n \sum_p (\gamma_{kp}^s + \lambda_{knp}^s) R_{knp}^s + \sum_j \theta_j^s FC_j + \sum_k \sigma_k^s DC_k ,
\end{aligned}
\tag{1}
$$

$s.t.$

$$\sum_j G_{ijm}^s \le cap_{im}^s X_{im}, \forall i, m, s , \tag{2}$$

$$\sum_j G_{ijm}^s \le cap_{im}^s X_{im}, \forall i, m, s , \tag{3}$$

$$\sum_j G_{ijm}^s \ge \sum_k \sum_p h_{mp}^s Q_{jkp}^s, \forall j, m, s , \tag{4}$$

$$\sum_j Q_{jkp}^s \ge \sum_n R_{knp}^s, \forall j, m, s , \tag{5}$$

$$\sum_k XY_{kn} = 1, \forall n , \tag{6}$$

$$R_{knp}^s \ge d_{np}^s XY_{kn}, \forall k, n, p, s , \tag{7}$$

$$\sum_k \sum_p a_{jp}^s Q_{jkp}^s \le FC_j, \forall j, s , \tag{8}$$

$$\sum_k \sum_p Q_{jkp}^s + \sum_n \sum_p d_{np}^s XY_{kn} \le DC_k, \forall k, s , \tag{9}$$

$$FC_j \le oY_j, \forall j , \tag{10}$$

$$DC_k \le oZ_k, \forall k , \tag{11}$$

$$\xi_s - \xi_s^* \le \omega \xi_s^*, \forall s , \tag{12}$$

$$G_{ijm}^s, Q_{jkp}^s, R_{knp}^s \ge 0, \forall i, j, k, m, n, p, s , \tag{13}$$

$$X_{im}, Y_j, Z_k, X_{kn} \in \{0,1\}, \forall i, j, k, n , \tag{14}$$

$$FC_j, DC_k \ge 0, \forall j, k . \tag{15}$$

Equation (12) ensure a feasible solution of objective function value in a given situation does not exceed the area of optimization objective function value. Equations (13)-(15) are the variable value range constraint.

If we do not consider Equation (12), the above model would be transformed into a stochastic programming model SPM Equation (1) *s.t.* Equations (2)-(11) and (13)-(15).

4 Algorithm design

4.1 FEASIBILITY OF SOLVING MODEL

When $\omega \to +\infty$, ROM is stochastic programming model actually, and constraint condition become redundant constraints in the model, cish8i has a feasible answer, through optimize algorithm can get best answer. But for regrets limited coefficient you need to do feasibility analysis to ROM.

Move Equation (12), then:
$\xi_s \le (1-\omega)\xi_s^*$. Let $UB = \sum_s \rho_s (1-\omega)\xi_s^*$, therefore we can get the theorem: If ROM works, UB is its upper bound. Using this theorem, we can get following ROM feasibility test rules:

For a given heuristic, if we do not take into account Equation (12), the target value $\xi > UB$, the algorithm cannot get optimized ROM solution, satisfactory solutions even a feasible solution. Otherwise, feasible solution can be obtained, and satisfactory solutions even optimal solution.

4.2 DETERMINISTIC OPTIMIZATION PROBLEMS

Before solve ROM, first study the given scenery, then determine algorithm for optimization NDs.

$$
\begin{aligned}
Min\xi = \sum_i \sum_m \varpi_{im}^s X_{im} + \sum_j v_j^s Y_j + \sum_k w_k^s Z_k + \\
\sum_i \sum_j \sum_m (\alpha_{im}^s + \mu_{ijm}^s) G_{ijm}^s + \sum_j \sum_k \sum_p (\beta_{jp}^s + \eta_{jkp}^s) Q_{jkp}^s + \\
\sum_k \sum_n \sum_p (\gamma_{kp}^s + \lambda_{knp}^s) R_{knp}^s + \sum_j \theta_j^s FC_j + \sum_k \sigma_k^s DC_k ,
\end{aligned}
$$

s.t. Equations (2)-(11) and (13)-(15).

This kind of problem is theoretically can obtain exact solution, due to large scale and the computation time so they are unacceptable by us.

4.3 ALGORITHM PROCESS

1) Initialize at first, then select 0-1 design variable's initial feasible solution x_{now}, Maximum number of iterations is *MN*, another number in the candidate collection *num*, initialization tabu list $TB = \phi$, tabu length $len = \sqrt{num}$. Make $bestsol_s = +\infty$, $step = 0$.

2) Choose several *num* from x_{now} neighbourhood to form candidate collection $can(x_{now})$.

3) According to the supply chain structure for each neighbourhood, using all or nothing heuristic algorithm for flow distribution.

4) Let $\xi(x_{nb}) = \xi_{CM}^s + \xi_{DM}^s, nb = 1, 2, ..., num$. Computing objective function values of each neighbourhood in the collection.

5) Determine whether the current iteration satisfies the amnesty rule, if no, selected optimal solutions x_{nb} that not be tabu in the candidate collection, according to the principle of first-in first-out to update tabu list, $x_{now} = x_{nb}$, otherwise, choose another feasible solution x_{nb} in candidate collection as tabu object $x_{now} = x_{nb}$.

6) If $bestsol_s > \xi(x_{now})$, record the present optimal solutions $bestsol_s > \xi(x_{now})$, $x_{best} = x_{now}$, $beststep = step$

7) If $step - beststep \le NM$, $step = step + 1$ then turn to the second item, if not, the algorithm finished.

5 Algorithm of ROM

To studying the feasibility of robust optimization model, we should study the stochastic programming model first.

ROM for studying the feasibility of the first to study the stochastic programming model. Actually, to obtain solution of stochastic programming model, it is only required to change the above 4th step of the algorithm: $\forall s \in S$. Let $\xi(x_{nb}) = \xi_{CM}^s + \xi_{DM}^s, nb = 1, 2, ..., num$, calculate $\xi(x_{nb}) = \sum_s \rho_s \xi_s(x_{nb})$.

After solving the stochastic programming model, according to feasibility tests, if there is a feasible solution, use the following steps to solve it.

5.1 INITIALIZE

Select 0-1 design variable's initial feasible solution x_{now}, given two target value has not changed maximum number of iterations *MN* and another neighbourhood number in the candidate collection *num*, initialization tabu list $TB = \phi$, tabu length $len = \sqrt{num}$. $RS_{best} = +\infty$, $step = 0$.

5.2 NODES' DEPLOYMENT

Choose several *num* from x_{now} neighbourhood to form candidate collection $can(x_{now})$.

5.3 FLOW DISTRIBUTION

According to the supply chain structure for each neighbourhood, use all or nothing heuristic algorithm for flow distribution.

5.4 UPDATED ROBUST OPTIMAL SOLUTION

$\forall s \in S$ let $\xi(x_{nb}) = \xi_{CM}^s + \xi_{DM}^s$, $nb = 1, 2, ..., num$, calculate $\xi(x_{nb}) = \sum_s \rho_s \xi_s(x_{nb})$, judge whether there is feasible solution x_{nb}, makes the scene objective function values to meet robust constraints $\xi_s(x_{nb}) \le (1+\omega)\xi_s^*$, if $\exists x_{nb}, RS_{best} > \xi(x_{nb})$, then $beststep = step$, record present optimal robust solutions $RS_{best} = \xi(x_{nb})$, $rx_{best} = x_{nb}$.

5.5 UPDATE TABU LIST

Determine whether the current iteration satisfies the amnesty rule, if not, elected optimal solution x_{nb} that not be tabu in the candidate collection, $x_{now} = x_{nb}$, otherwise, choose another feasible solution x_{nb} in candidate collection as tabu object $x_{now} = x_{nb}$.

5.6 TERMINATION CONDITIONS OF ALGORITHM

If $step - beststep \le NM$, $step = step + 1$, then turn to second item, or this algorithm finished.

6 Application of ROM

Assume that the characters of study object are: $|N| = 30, |I| = 3, |J| = 3, |K| = 3, |M| = 6, |P| = 6$, scenery number is 10. Use robust optimization model and stochastic programming model to do supply chain network design, and then compare performance of optimized solution that is obtained.

Make every scenery's deterministic optimal objective function value is ξ_s^*, stochastic programming optimization objective function value corresponding to each scenery as ξ_s^{SP}, robust optimal solution corresponding to each target value for ξ_s^{RO}.

Take the relative difference value between ξ_s^{SP} and ξ_s^*: $\varepsilon_{S-D} = \frac{\xi_s^{SP}\xi_s^*}{\xi_s^*} \times 100\%$.

Relative difference value between ξ_s^{RO} and ξ_s^*:

$$\varepsilon_{R-D} = \frac{\xi_s^{RO}\,\xi_s^*}{\xi_s^*}\times 100\%\ .$$

Computed result:

TABLE 2 Comparison chart of robust optimization and stochastic programming solutions' performance

Stochastic programming ξ_s^{SP}	Robust optimization ξ_s^{RO}	Optimal objective values ξ_s^*	Relative difference value ε_{S-D}	Relative difference value ε_{R-D}
5589027	565680	558474	0.1	1.29

Using the robust optimization model for design, system cost (the objective function value) for different scenarios, and if fluctuations are relatively stable, rendered as insensitive, implies that this determined supply chain network structure has less market risk. As a strategic supply chain network design, not only to establish the basis for subsequent development of enterprises, also involves saving a lot of fixed assets foe investment cost, of course, is also very necessary to reduce risks.

7 Conclusion

This chapter studied strategy-level of integration supply chain network design problem under uncertain environment, the aim is to take into account the uncertainty of design parameters, determine the selection of suppliers, factories and distributors' locations as well as the capacity, customer demand, and to make performance in supply chain networks have a good robust in the case of parameter perturbation. Through detailed analysis of the problem of supply chain network design, we established the integration of supply chain network design for robust optimization model and stochastic programming models.

In this algorithm design process, first of all, we analyses the feasibility of solving the model, and then to study the deterministic optimization problems. Proposed robust optimization model algorithm based on deterministic optimization model algorithm.

In the last example, we comparing difference in solving the objective function value between stochastic programming solutions and robust optimization solutions, evaluated performance of supply chain network that confirmed by robust optimization models. Calculations results not only show that the all or nothing algorithm of tabu search has good convergence properties, as well as its advantages in dealing with large scale problems, but also show that the robust optimization model for supply chain network design can effectively reduce market risk, which is essential for decision-making at the strategic level.

References

[1] Junfeng T 2005 Supply chain management optimization models and Algorithms under uncertainty PhD *Thesis of Southwest Jiaotong University (in Chinese)*

[2] New S J New 1996 A framework for analysing supply chain improvement *International Journal of Operations & Production Management* 16(4) 19-34

[3] Itoh T, Ishii H 2005 One Machine Scheduling Problem with Fuzzy Random Due-Dates *Fuzzy Optimization and Decision Making* **4**(1) 71-8

[4] Cao J, Zhou G G, Zhang D Y 2006 A Multi-plan synthetic evaluation method based upom AHP and fuzzy theory *Journal of Zhejiang University of Technology* **31**(4) 355-9 *(in Chinese)*

[5] Jjin X, Zhenfa Z 2004 An Analysis of Supplier's Competence Based on ANN Expert System *Industrial Engineering and Management* **2** 40-3

[6] Taylor (USA) 2006 Success to supply chain: Supply chain competition determine the success *China Market Press*

[7] Ferguson R B 2004 Modules cut supply chain risk *e-Week* **21**(3) 24

[8] Junyou J 2005 Analysis of Supply Chain Risk *Baocheng Technology* **2** 35-8 *(in Chinese)*

[9] Zhi L, Tao Y, Wen Y 2006 Supply chain risk analysis and decision-making model *Logistics Technology* **29**(130) 120-2 *(in Chinese)*

[10] Xiaoyu Z, Xiaoyuan H, Fuquan S 2005 Supply chain design method toward key suppliers and customers *Systems engineering* **23**(6) 34-8 *(in Chinese)*

[11] Nagurney A, Matsypura D 2005 Global supply chain network dynamics with multicriteria decision making under risk and uncertainty *Transportation Research Part E* **41**(6) 585-612

[12] Satyaveer S C, Jean M P 2005 Analysis of supply chain partnership with revenue sharing *International Journal of Production Economics* **97**(1) 44-51

[13] Jun P 2008 Design and Application of Supply Chain Optimization consider risk control *Hunan University (in Chinese)*

Author

Yulan Zheng, Fuzhou Province of China

Current position, grades: lecturer.
University studies: PhD degree will be earned in major of technology economy and management, Fuzhou University.
Scientific interest: technological innovation and management.

An empirical study on China listing corporation industrial-financial combination based on the regression method

Zhiming Wang[1, 2*], Shuzhen Zhu[1]

[1]*Glorious Sun School of Business and Management, Donghua University, Shanghai, 200051, P R China*

[2]*Shandong Iron and Steel Co., Ltd. Laiwu Branch, Laiwu, 271104, P R China*

Abstract

This paper takes Chinese listing corporation as research samples to find the best combination area or point about the industry-finance combination on enterprise operating performance. Research on the combination of behaviour impact is carried out through EPS, RPE, ROE, Tobin Q four indicators. The paper is mainly using the nonlinear regression method based on steel listing corporation annual data from 2005 to 2008. Experiments show that the relation by EPS, RPE and Tobin Q meeting the cubic function curve is significant; the relation by ROE meeting the curve function is not significant. In a certain stage of development of the enterprise the appropriate industry-finance combination area or ratio exists, but the ratio of each index is different.

Keywords: industrial-financial combination, the nonlinear regression, the cubic curve, China steel listing corporation, the best combination point

1 Introduction

The combination of industry and finance is the industrial economy and the financial sector in the development process of the mutual penetration and influence process. It is that the market economy develops to a certain stage of the inevitable trend [1, 2].

Practice of the developed countries shows that industrial capital and financial capital will have a fusion process to distribute the social resources more effective to meet the objective requirements. This kind of fusion is conducive to the national financial policy at the macro level and fast flowing to industrial capital at the micro level, and can ultimately improve the efficiency of capital allocation. Industry and finance how to combine effectively became one of the research foci of the theory. Many scholars have studied the problem and the concept of the integration essence, mechanism, motivation and the effective conditions [3, 4].

In recent years, Baosteel Group, Shougang Group, Wuhan Iron and Steel Group, China steel enterprises listed company have carried out the exploration and the practice of the industry and finance combination accumulated certain experience. These practice and explorations as we stand on the corporate perspective of the Combination of behaviour provide realistic materials.

This article takes Chinese listing corporation as research samples to find the best combination area or point about the industry-finance combination on enterprise operating performance, making an empirical test of different proportion of share of the combination of industry with finance behaviour influence on the management performance and the value of the enterprise.

Research on the combination of behaviour impact is carried out through EPS, RPE, ROE, Tobin Q four indexes. In order to obtain a smooth curve fitting for scattered data to find the best combination points or area is the research focus and difficulty of microscopic study. This paper is mainly using the nonlinear regression method based on China's steel listing corporation annual data from 2005 to 2008. Experiments show that the relation by EPS, RPE and Tobin Q meeting the cubic function curve is significant; the relation by ROE meeting the curve function is not significant. This text fully shows that the combination optimum ratio do exist in a certain stage of development of companies.

2 Literature review

The existing literature about the combination effect of listing Corporation empirical research mainly concentrated in two observation aspects. One is based on financial enterprises shares of non-financial listing corporation perspective, using cross section data in quantitative analysis of finance enterprise shareholding proportion of listing corporation performance and other aspects of the impact effect of listing Corporation empirical research.

Another is based on non-financial listing corporation shares of financial institutions, contrast before and after financial listing corporation accounting index to analyse the impact of combination of industry with finance. This study is based on the first one [5, 6].

We have found that the enterprise in a certain stage of development has the most appropriate industry-finance combination ratio or area to make the impacts on the

---

* *Corresponding author* e-mail: lgyj0824@163.com

performance and the value by using Loess regression method in 2013.
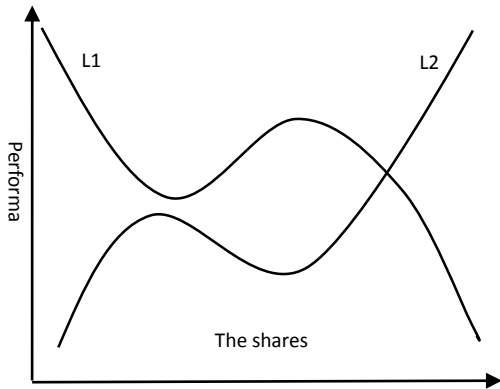
FIGURE 1 The Cubic curve of the relation between the performance and the shares

Further documents showed, the study is essentially research on the relationship between firm performances and the larger shareholders holdings based on the viewing angle. Existing literatures found that the largest shareholder of the company effect on the value may not only bring the potential incentive effects, but also bring defensive effects. The impact is with a clear section feature. There is the Cubic regression relationship between corporate value and the first shareholder ownership. But the change trend of the curve of the first major shareholders is opposite to that of the second largest shareholder. As Figure 1, curve L1 shows the relationship between corporate performance and the proportion of the first largest shareholder, curve L2 shows the relations between corporate performance and the proportion of the second largest shareholder [7, 8].

3 Data and methodology

3.1 VARIABLES AND DATA

With reference to the relevant literature we selected RPE, EPS, ROE, Tobin Q to represent the enterprise performance.

The Scale variable, used to describe the degree of industry-finance combination, is the total of the proportion held by shareholders with financial background from the 10 larger shareholders by our collection and arrangement.

This paper studied Chinese Shanghai and Shenzhen Stock Markets Sample steel listing Corporation (C6), with the time interval from 2005 to 2008. In this paper, data from Shanghai and Shenzhen stock market listing Corporations released annual report of listing Corporation *.pdf file or bulletin. The main data are from RESSET database, CHINF sites, China Securities Regulatory Commission and China listing Corporation information network and other related sites.

Considering the effectiveness and availability of data, financial shareholder equity ratio of the top ten shareholders is selected as the explanatory variable.

Those heterogeneous samples are eliminated. For example, the same company's stock has been listed in Hong Kong stock market, or EPS is negative or more than 100, or ROE is negative, or the stock symbol is ST.

Through screening and analysis, we derive conditions consistent with the sample of 95 listing Corporation. The data used variables are defined as shown in Table 1. Results are analysed and calculated by using SPSS 18.

As shown in Table 2, RPE, EPS and ROE of the average of 18.03, 0.376, 0.184 and Tobin Q value of mean 1.376, mean that iron and steel enterprises are still in rapid development, and benefits is good, and market prospect is valued and recognized from 2005 to 2008. In fact, China steel industry is in the best period of its history.

3.2 The nonlinear regression method

Firstly we study variables Pearson correlation analysis. Then according to the degree of correlation coefficient, the regression models are established. Finally, the function models are utilized to analyse the relevant data.

In the light of the previous literature, we put the proportion of shares of financial companies as an explanatory variable, and selected RPE, EPS, ROE and Tobin Q value as the explained variables. So four regression models are established [5, 6].

According to the trend of correlation analysis and plot, we use the SPSS software regression analysis, curve estimation, regression analysis of data using the linear, logarithmic, countdown, two times, three times, compound, power, S, growth, index, Logistic function model.

TABLE 1 The definition of variables

Variable	Symbol	Data type	Definition
Order	no	Numeric(N)	The order
Company code	code	Character(C)	Steel listed companies code
Company name	company	Character(C)	Steel listed companies name
Data date	date	Date(D)	The date of the data
The shareholding ratio	scale	Numeric(N)	The total of the proportion held by shareholders with financial background from the 10 larger shareholders
P/ e ratio	rpe	Numeric(N)	The annual price-earnings ratio
Earnings per share	eps	Numeric(N)	Diluted earnings per share
The net assets returns ratio	roe	Numeric(N)	Diluted net assets returns ratio
Tobin Q	qval	Numeric(N)	The Tobin's value
Remark	bj	Character(C)	Notes and instructions

TABLE 2 The descriptive statistics of variables

Symbol	N	Min	Max	Mean	Std. Deviation
Date	95	2005/12/31	2008/12/31		
Scale	95	0.15	39.28	6.2767	7.0096
Rpe	95	**3.9332**	**90.6077**	18.0323	16.71
Eps	95	0.0061	1.2356	0.3756	0.3027
Roe	95	0.0033	0.88	0.184	0.1815
Qval	95	0.7936	3.0421	1.3257	0.554

TABLE 3 The Pearson correlations of variables

		Scale	RPE	EPS	ROE	Qval
Scale	Pearson Correlation	1	.240*	.285**	-.029	.316**
	Sig. (2-tailed)		.019	.005	.779	.002
	N	95	95	95	95	95

* Correlation is significant at the 0.05 level (2-tailed) **. Correlation is significant at the 0.01 level (2-tailed)

As can be seen from Table 3, between EPS and scale, the Pearson correlation coefficient was 0.285, and at the 0.01 level significantly. So there is an obvious positive correlation between the two variables. The function of curve fitting regression analysis by SPSS software, the results in Table 4 and Figure 2.

TABLE 4 The model summary and parameter estimates of EPS

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.081	8.252	1	93	.005	.298	.012		
Logarithmic	.092	9.464	1	93	.003	.271	.080		
Inverse	.026	2.531	1	93	.115	.403	-.045		
Quadratic	.146	7.836	2	92	.001	.197	.041	-.001	
Cubic	.147	5.217	3	91	.002	.210	.034	.000	-1.456E-5
Compound	.072	7.269	1	93	.008	.192	1.041		
Power	.104	10.840	1	93	.001	.169	.292		
S	.053	5.182	1	93	.025	-1.265	-.217		
Growth	.072	7.269	1	93	.008	-1.648	.040		
Exponential	.072	7.269	1	93	.008	.192	.040		
Logistic	.072	7.269	1	93	.008	5.197	.961		

The dependent variable is EPS, and the independent variable is scale.

Table 4 shows the three function model (Cubic) fitting degree is higher, and the overall model fit is higher. We can get the equation model as follows. The h in Equation (1) represents the scale variable.

$$EPS = c + \beta_1 h + \beta_2 h^2 + \beta_3 h^3 . \tag{1}$$

The curve of EPS is as Figure 2.

TABLE 5 The cubic model summary of RPE, ROE and Qval

Dependent Variable	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
RPE	.066	2.149	3	91	.099	14.820	.066	.073	-.002
ROE	.021	.656	3	91	.581	.235	-.021	.002	-3.145E-5
Qval	.145	5.130	3	91	.003	.972	.093	-.004	5.428E-5

Seen from Table 5, RPE and Qval (Tobin Q) reach statistical significant criteria. ROE does not reach the significant requirements. The h in equations represents the scale variable.

$$RPE = c + \beta_1 h + \beta_2 h^2 + \beta_3 h^3 , \tag{2}$$

$$Tobin\,Q = C + \beta_1 h + \beta_2 h^2 + \beta_3 h^2 . \tag{3}$$

The Cubic curve of the variables is as follows figures.

In order to observe the overall change trend, we will coordinate axes appropriately extended. It is no practical significance that some coordinate is negative. For example, x is negative of no significance
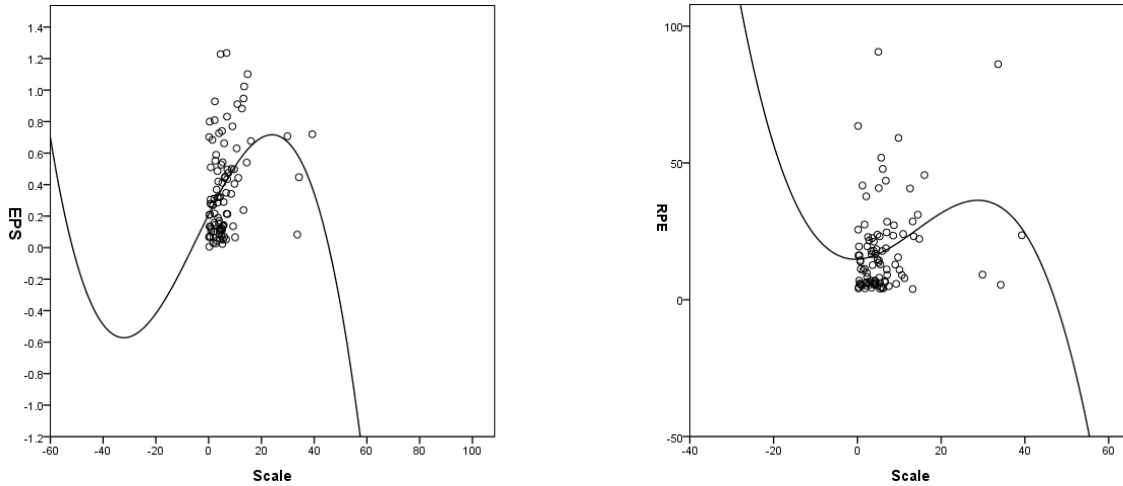
FIGURE 2 The Cubic curve of EPS and RPE
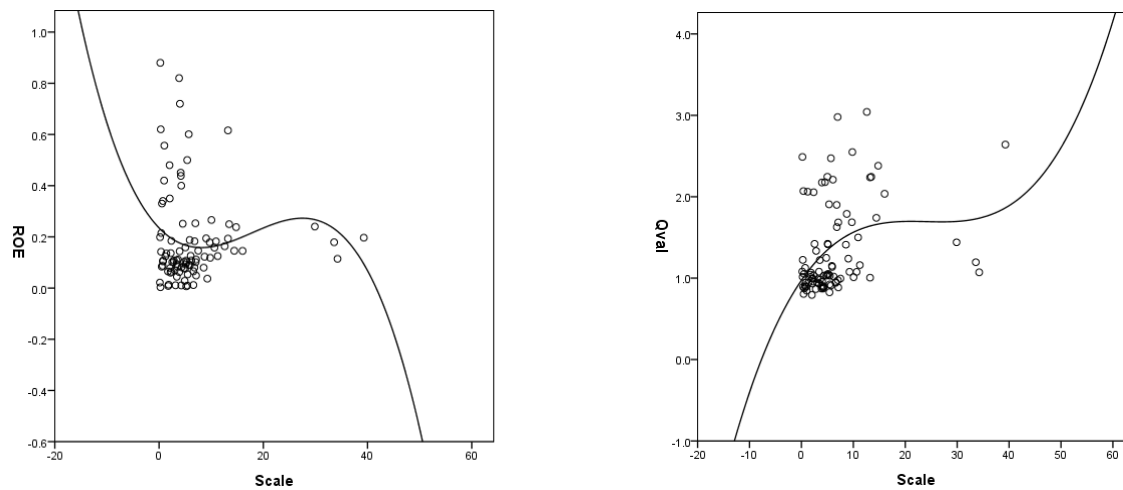
FIGURE 3 The Cubic curve of ROE and Tobin Q

4 Empirical Results

Combined with the above, we can see the following phenomena as follows:

First, from Figure 2 to Figure 3 can be seen, the curve is a "S" type, and exists the extreme values in certain interval. According to the equation each variable can calculate x coordinates of each variable in the range of the maximum or minimum value. For example, when x value is 27.9, EPS reaches the maximum value in the right interval. When x value is 24.8, RPE reaches the maximum value in its right interval. And when x value is 6.1 or 36.3, ROE can reach the maximum value in its interval. And when x value is 18.9 or 30.2, Tobin Q can reach the maximum value in its interval. EPS and RPE have one value, but ROE and Tobin Q have two values. This is consistent with the figures as Figure 2 and Figure 3.

Second, RPE, EPS, ROE and Tobin Q can reach the maximum value or the minimum value from the Cubic curves, but the ratio of each index is different.

In the end, from 2005 to 2008 by our calculation, the means of the combination of ownership is increased gradually, and the maximum value is 8.41 in 2007.

Because of 2008 financial crisis consequences, the ownership began to decline significantly in 2008.

5 Conclusions and prospect

Three important conclusions can be drawn from the above results.

Firstly, EPS, RPE and Tobin Q could obey the rules of shareholding ratio and operating performance. But the law obeyed by EPS and RPE is different from the one obeyed by Tobin Q. ROE does not reach statistical significance criterion.

Secondly, the combination optimum ratio about the relation between the performance and the shares do exist, but the ratio of each index is different. The combination has different effects on each index of the iron and steel enterprises. RPE, EPS and Tobin Q can reach the maximum value or the minimum value when the proportion is appropriate, but the value of each index is different.

Finally, as the ratio is far from the optimal proportion for the maximum value of the mean scale is only 8.41, the industry-finance combination of Chinese iron and steel listing Corporation is still relatively junior.

Chinese iron and steel enterprises should consider the combination of industry and finance as the priority direction of development as the steel industry is a typical capital-intensive, resource-intensive and technology-intensive industry. It is necessary for companies to suit their measures to local conditions to carry out the combination practice. Only when certain conditions would be met, production and finance can be effectively combined [9]. To speed up the upgrading of the industrial structure adjustment, the combination provides a new way and an idea for China iron and steel industry and other traditional industries. It can promote China's iron and steel enterprises to get out of the current predicament of production and management as soon as possible [10, 11].

Acknowledgements

References

[1] Laforge 1985 Selected works of Laforge *People's Publishing House Press*2-51 (*in Chinese*)

[2] Stiroh K J, Rumble A 2006 The Dark Side of Diversification: The Case of US Financial Holding Companies *Journal of Banking & Finance* **30** 2131-61

[3] Fotios Pasiouras 2008 Estimating the Technical and Scale Efficiency of Greek Commercial Banks: The Impact of Credit Risk, Off - balance Sheet Activities and International Operations *Research in International Business and Finance* **22**(3) 301-18

[4] Fu Yan 2003 The IIFC road to where, Chinese industry and study of the effectiveness of financial integration *People's publishing house* 11-59 (*in Chinese*)

[5] Zhang Q, Yang L 2005 The industry-finance group: development practice in foreign countries and China *China Financial Publishing House* 2-79 (*in Chinese*)

[6] Yao D, Wang S ,Luo C et al. 2011 The empirical evaluation of the operational efficiency of the Combination of listed companies *China Soft Science* **3** 140-8 (*in Chinese*)

[7] Xiong X-z,Li S-m, Li J 2008 Outside Blockholders, Tunneling Behavior and Corporate Performance: An Empirical Evidence from China's IPO Firms *Systems Engineering* **26**(5) 17-21(*in Chinese*)

[8] Xie J 2007 The Largest Shareholding and Corporate Valuation: Incentive Effect and Entrenchment Effect. *Nankai Business Review* **10**(1) 21-5 (*in Chinese*)

[9] Yin R 2004 The rise of China's iron and steel industry and technology in progress *Metallurgical Industry Press* 5-82 (*in Chinese*)

[10]Qian Z X 2010 Industry finance *Jiangsu People's Publishing House* 56-81 (*in Chinese*)

[11]Hu Y 2008 Economics of industry convergence analysis *China Economic Publishing House* 26-63 (*in Chinese*)

Author

Zhiming Wang, born on April 21, 1973, Dongming, China

Current position: engineer of Chandong Iron and Steel Co., Ltd. Laiwu Branch, P.R. China.
University studies: Ph.D. student of Glorious Sun School of Business and Management, Donghua University, China.
Scientific interest: asset management, financial management, financial modelling.
Publications: 10 papers.
Experience: about 20 years.

Shuzhen Zhu, born on October 18, 1960, Shanghai, China

Current position: Doctor of Management science and engineering, professor and doctoral supervisor in Donghua University.
University studies: Management science and engineering in Donghua University.
Scientific interest: risk management, financial innovation, finance engineering.
Publications: 2 Patents, 50 Papers.
Experience: about 30 years.

Research on grey correlation analysis model of enterprise human resources competitiveness

Xiuli Li[1*], Yuhong Zhang[2], Sujuan Zhao[1]

[1]*College of finance, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P.R. China*

[2]*College of education, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei, P.R. China*

Abstract

Human resources is one of the key indexes in measuring the competitiveness of en enterprise. However, evaluation of enterprise human resource competitiveness, evaluation index system and evaluation model remain a problem in academic and practice. Nowadays, key factors in evaluation of enterprise human resources competitiveness and evaluation index are incomplete and imperfect, and processing methods of corresponding evaluation index is not scientific. Thus, studies on analysis model of enterprise human resource are of great importance. In this paper, an improved grey correlation analysis model of human resource competitiveness was put forward. This model provides relative restrictive factors in analysis of enterprise resource competitiveness, and analysed enterprise competitiveness from the aspects of human resource quality competiveness, market development competiveness, management quality competiveness, etc. and evaluation analysis based on improved grey correlation analysis method was conducted. Its evaluation result can be a basis of selection of human resource developing strategies for direction and of frame of decision-making. Finally, the model and algorithm was proved feasible by implementation of actual case.

Keywords: human resources, competitiveness, grey correlation analysis, evaluating indicator, model

1 Introduction

Enterprise human resources refer to the human resources or labor force owned or controlled by the enterprise that can bring economic interest in the process of production and management. Along with the development of information technology and computer science and technology, the importance of enterprise human resource is increasing. And the emphasis of enterprise human resources not only on human resources or labour force, but has expanded to the strength and quality an enterprise owned that can make the enterprise obtain economic profit and develop sustainably, which embodies various aspects including the management ability of human resources, human resource competitiveness and extensible developing potential based on science and technology [1-3]. Thus, analysis of human resource evaluation needs to be based on the aspects of human resources, including sociality, reproducibility, profitability, marketability, dynamics, hierarchy and subjective initiative, for comprehensive analysis, and obtain key evaluation factors of its evaluation analysis of human resource competitiveness. And comprehensive evaluation index system of enterprise human resource competitiveness can be created, which will provide strong support for implementation of computer aided intelligent analysis method of evaluation of enterprise human resource competitiveness and implementation of the system [4-5]. By far, some relative studies had analysed and probe into

enterprise human resources, and had provided good directions and strategies of implementation [6-10]. However, most of the existing studies on analysis of human resource competitiveness limited to provide relative guiding strategies based on quantitative analysis, cannot implement comprehensive evaluation analysis combining qualitative and quantitative aspects, and cannot provide qualitative analysis models, which can guide the analysis of enterprise human resources. Thus, this paper, based on existing researches and studies, via analysis of relative restrictive factors that influence enterprises human resources, established a comprehensive evaluation index system of enterprise human resources corresponding to restrictive factors, and established corresponding comprehensive evaluation calculation model of enterprise human resources based on grey correlation analysis.

2 Influence factors in enterprise human resource and comprehensive index system

2.1 ANALYSIS OF INFLUENCE FACTORS IN ENTERPRISE HUMAN RESOURCE

Restraint of social factors: human resource in a certain population engages in social labour. It cannot leave the society, and is limited by various conditions including social politics, economy and culture. Thus, human resource presents strong sociality. Sociality is the essential attribute of human resource, and is also the basic difference between human resource and material resource.

---

[*] *Corresponding author* e-mail: 32590984@qq.com

It needs to be noted that along with the development of society and science and technology, human resources does not refer to human, but the individual capacity carried by human. That is to say, the carrier of human resource is human, and the core of human resources is labour force.

Restraint of reproducibility factors: reproducibility of human resources includes reproducibility of enterprise labour force and reproducibility of human resources. The reproducibility is realized via the process of continuous replacement, renewal and rehabilitation of the entire labour force population of enterprise and each individual in human resources. Along with the replacement of life circle of product designation and the development of social science and technology, the production technology of enterprise labour force will be supplemented and renewed constantly for adjusting to the requirements of designation, renewal and maintaining of new products. Thus, for improving of enterprise labour productivity and realizing reproducibility of enterprise human resources, influences and restraints of various reproducibility factors need to be.

Restraints of profitability factors: by combining with other kinds of resources, enterprise human resources can bring presumptive increment of value. But it needs to be noted that human resource is different from other forms of tangible resources. The enterprise human resource is not only restraint by internal factors of individual labour, but also influenced by organization management within the enterprise, economic environment out of the enterprise and other constraints, which makes its profitability full of uncertainty.

Restraint of market factors: enterprise human resource is an activity that serves business market. When in market activities implement, value evaluation of human resource obtain the evaluation result via the forming process of asset value and mechanism of action under the condition of simulative market. Thus, development potential and prospect of market is the key segment in enterprise human resource project. The enterprise human resource needs to correlate closely with the development trend and potential development competitiveness to combine effectively human resources, material resources and intangible assets in the process of market prediction and market exploitation and cannot deviate off market development. Thus, it can bring economic benefit to the enterprise.

Restraints of dynamic factors: dynamic refers to that value of human resource is evaluated from dynamic perspective. First of all, because human resource and its carrier cannot be separated, human resources cannot be quantized with money in static state like material resource. It can only be evaluated in dynamic state, namely the performance appraisal in its using process. The performance is not only decided by its own value, but also by environment, organizational system, supply and demand, etc. These variables of the value of human resources change constantly, which makes the evaluation of the value of human resources full of uncertainty.

Restraint of hierarchy: human resource is a new capital form, which is put forward directing at assumption of capital homogeneity in traditional theory. As a breakthrough of capital homogeneity, within human resources there is obvious heterogeneity，i.e. hierarchy. That is to say, human resource is influenced by the factors of hierarchy including learning capacity, working ability, creativity, ability to organize and manage and resources allocation abilities of the carrier of labour force.

Restraints of subjective initiative: the subjective initiative of human resource is a key segment industry in obtaining innovative development. That is to say, the carrier of human resources can conduct creative activities with certain purposes, and can change the world via its own labour. All the economic activities show up firstly as activities of human resources. Activities of human resources occupy the most important status in economic activities. Human resource is a very active economic resource and productive factors, which launches, organizes and coordinates other resources, and is the only factor, which can play the role of creativity in economic activities. The subjective initiative and creativity of human resource plays a decisive role in the development of social production.

2.2 ESTABLISHMENT OF COMPREHENSIVE EVALUATION INDEX SYSTEM OF ENTERPRISE HUMAN RESOURCE COMPETITIVENESS

To establish the corresponding comprehensive evaluation index system of enterprise human resource competitiveness, the human resource competitiveness of the whole enterprise should be evaluated from comprehensive perspectives both inside and outside of enterprise. And the influence of restraints mentioned above need to be taken into consideration, and analysis from the level of three constraints, namely human quality competitiveness, market competitiveness and management quality competitiveness, needs to be done. What's more, the evaluation indexes should be scientific, complete, comparable and operable. The evaluation index system is presented in Figure 1.

Target layer   Factor level   Index layer
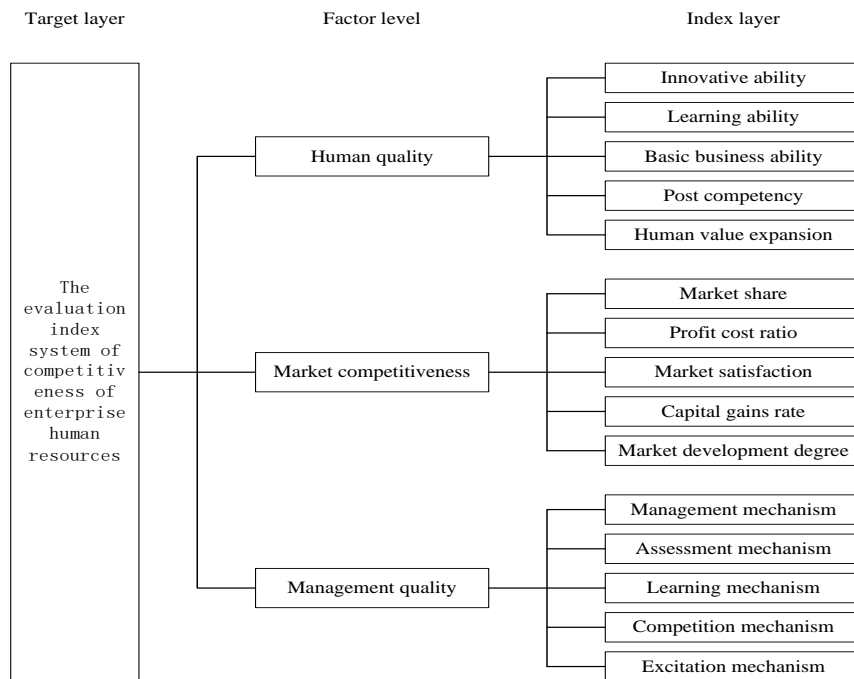
FIGURE1 The comprehensive evaluation index system of enterprise human resource competitiveness

The human quality competitiveness refers to technology and ability of enterprise employees. The ability and technology enables them to be qualified to their positions. The index of human quality competitiveness manifests as evaluation of human resource competitiveness from inside of enterprises. This category of index can comprehensively evaluate the capability of employees. Market competitiveness refers to the ability of an enterprise to occupy and expand the market. The index of market competitiveness is one of the important indexes that reflect the human resource management ability of an enterprise. And this category of indexed is outcome indexed that influence human resource competitiveness. They can work as objective index factors in adjustment of framework of human resource competitiveness which reflects the prospect of human resource management and make the human resource development strategic. Management quality competitiveness refers to rules related to personal management including system and rules of employment. Rational management, planning and employment cannot only improve human resource competitiveness but also promote the innovation and development of enterprises.

3 Grey correlation analysis model of enterprise human resources competitiveness

3.1 STANDARDIZATION OF EVALUATION INDEXES

Under the condition of that the comprehensive evaluation index system of enterprise human resource competitiveness has been built, evaluation indexes in different categories need to be standardized. According to

the analysis of evaluation indexes mentioned above, evaluation indexes influenced by different factors have been transformed as positive indexes. Thus, in this paper, this paper applied ratio scale from 1 to 9 to mark fuzzy grade to evaluation indexes in different categories. Thus, fuzzy evaluation values corresponding to different types of evaluation indexes can be obtained. The concrete evaluation standard is presented in Table 1.

TABLE 1 The evaluation criteria of the standardization of evaluation indexes

Ratio scale	Importance degree	Explanation
1	Not important	Of the least importance
3	Fairly important	Of fairly importance for evaluation analysis
5	Important	Of mediate importance for evaluation analysis
7	Vary important	Of considerate importance for evaluation analysis
9	Absolutely important	Of the highest importance
2,4,6,8	Median between two neighbouring levels	Applied when an average needs to be stroke

3.2 GREY CORRELATON ANALYSIS OF EVALUATION INDEX

Grey correlation analysis is a correlation measuring method that sorts system characteristics and relative factors according to the corresponding calculated grey correlation. It is an analysis method that combines qualitative method and quantitative method, which can successfully solve the problem in evaluation indexes quantization and analysis. It can also make the evaluation result more scientific, objective and accurate by avoiding influence taken by human factors [11-14]. According to the theory of grey correlation, the closer the geometrical shapes of curves of two sequences, the larger the

341

correlation between them. Thus, after obtaining the fuzzy values of different indexes $\gamma_{ij} = \left[\gamma_{ij}^{lef}, \gamma_{ij}^{rig} \right]$, $\gamma_{ij}^{lef} \leq \gamma_{ij}^{rig}$, positive ideal grey correlation sequence of evaluation indexes can be built as:

$$\gamma_{i0}^{R} = \left[\gamma_{i0}^{R-lef}, \gamma_{i0}^{R-rig} \right] = \max \left[\left(\gamma_{ij}^{lef} \mid 1 \leq j \leq m \right), \max \left(\gamma_{ij}^{rig} \mid 1 \leq j \leq m \right) \right]. \tag{1}$$

Similarly, negative ideal grey correlation sequence of evaluation indexes can be built as:

$$\gamma_{i0}^{N} = \left[\gamma_{i0}^{N-lef}, \gamma_{i0}^{N-rig} \right] = \left[\min \left(\gamma_{ij}^{lef} \mid 1 \leq j \leq m \right), \min \left(\gamma_{ij}^{rig} \mid 1 \leq j \leq m \right) \right]. \tag{2}$$

In the Equations (1), (2) i is the parameter and j is the number of projects under implementation.

Thus, the Euclidean distance between the value of evaluation index $\gamma_{ij} = \left[\gamma_{ij}^{lef}, \gamma_{ij}^{rig} \right]$ and the positive ideal grey correlation sequence is:

$$d_{i}^{R} = \sqrt{\frac{| \gamma_{i0}^{R-lef} - \gamma_{ij}^{lef} | + | \gamma_{ij}^{rig} - \gamma_{j0}^{R-rig} |^{2}}{2}}. \tag{3}$$

Thus, the Euclidean distance between the value of evaluation index $\gamma_{ij} = \left[\gamma_{ij}^{lef}, \gamma_{ij}^{rig} \right]$ and the negative ideal grey correlation sequence is:

$$d_{i}^{N} = \sqrt{\frac{| \gamma_{i0}^{N-lef} - \gamma_{ij}^{lef} | + | \gamma_{ij}^{rig} - \gamma_{j0}^{N-rig} |^{2}}{2}}. \tag{4}$$

Thus, the grey correlation index δ_{ij}^{R} of the human resource implementation plan P_j and positive ideal grey correlation sequence of the index γ_{ij} is:

$$\delta_{ij}^{R} = \frac{\min_{i} \min_{j} d_{i}^{R} + \beta \max_{i} \max_{j} d_{i}^{R}}{d_{i}^{R} + \beta \max_{i} \max_{j} d_{i}^{R}}. \tag{5}$$

In the Equation, β is resolution ratio，usually taken as $\beta = 0.5$.

The grey correlation index δ_{ij}^{N} of the human resource implementation plan P_j and negative ideal grey correlation sequence of the index γ_{ij} is:

$$\delta_{ij}^{N} = \frac{\min_{i} \min_{j} d_{i}^{N} + \beta \max_{i} \max_{j} d_{i}^{N}}{d_{i}^{N} + \beta \max_{i} \max_{j} d_{i}^{N}}. \tag{6}$$

The weighted grey correlation k_{j}^{R} of human resource implementation plan P_j and the positive ideal grey correlation sequence:

$$k_{j}^{R} = \sum_{i=1}^{n} \left(w_i \; \delta_{ij}^{R} \right). \tag{7}$$

The weighted grey correlation k_{j}^{N} of human resource implementation plan P_j and the positive ideal grey correlation sequence:

$$k_{j}^{N} = \sum_{i=1}^{n} \left(w_i \; \delta_{ij}^{N} \right). \tag{8}$$

3.3 GREY CORRELATION ANALYSIS MODEL OF HUMAN RESOURCE COMPETITIVENESS AND ALGORITHM IMPLEMENTATION

Human resource implementation plan P_j is attached to the optimal implementation plan. The membership of positive ideal grey correlation sequence is $\varphi_j \left(0 \leq \varphi_j \leq 1 \right)$, namely the grey correlation. And the membership of it attached to negative ideal grey correlation sequence is $1 - \varphi_j$. In order to determine the comprehensive grey correlation φ_j, establish the objective function according to the strength and weakness of the implementation plan:

$$F \left(\varphi_j \right) = \min \left\{ \left(\varphi_j \, k_{j}^{R} \right)^{2} + \left(\left(1 - \varphi_j \right) k_{j}^{N} \right)^{2} \right\}. \tag{9}$$

Obtain the correlation of grey correlation φ_j of the human resource implementation plan P_j according to extremum principle:

$$\varphi_j = 1 / \left(1 + \left(k_{j}^{N} / k_{j}^{R} \right)^{2} \right), \; j = 1, 2, L, m. \tag{10}$$

According to the above-mentioned grey correlation decision analysis, the grey correlation φ_j of each implementation plan can be obtained. An obtain the principle of proximity of human resource competitiveness grey correlation analysis of multi-attributes based on the grey correlation φ_j .If

$$\varphi_k = \max \left\{ \varphi_1, \varphi_2, L, \varphi_m \right\}. \tag{11}$$

Then human resource implementation plan P_k is the closest to the ideal grey matter-element, namely the human resource implementation plan P_k is the optimal implementation plan.

4 Case analysis

In this paper, the human resource competitiveness evaluation of a certain enterprise was analysed based on the abovementioned algorithm and model. Via consulting relative specialists, divide its human resource

competitiveness into four grades, namely "good", "fairly good", "ordinary" and "bad". Table 2 presents fuzzy

values of concrete evaluation indexes of the enterprise.

TABLE 2 The fuzzy value of enterprise human resource competitiveness

Influence factor	weight	Evaluation index	weight	Evaluation grade			
				Fairly good	Good	Ordinary	Bad
Human resource competitiveness	0.30	Innovative ability	0.20	8.0-9.0	7.0-8.0	6.0-7.0	3.0-5.0
		Learning ability	0.20	7.0-8.0	8.0-9.0	6.0-7.0	3.0-5.0
		Professional ability	0.25	6.0-7.0	7.0-8.0	3.0-5.0	8.0-9.0
		Basic business ability	0.20	6.0-7.0	8.0-9.0	7.0-8.0	3.0-5.0
		Human value expansion	0.15	8.0-9.0	6.0-7.0	7.0-8.0	3.0-5.0
Market competitiveness	0.30	Market share	0.25	3.0-5.0	6.0-7.0	8.0-9.0	7.0-8.0
		Profit cost ratio	0.15	8.0-9.0	7.0-8.0	3.0-5.0	6.0-7.0
		Market satisfaction	0.20	6.0-7.0	8.0-9.0	6.0-7.0	3.0-5.0
		Profit gains rate	0.15	7.0-8.0	6.0-7.0	8.0-9.0	3.0-5.0
		Market development degree	0.25	7.0-8.0	3.0-5.0	8.0-9.0	6.0-7.0
Management quality competitiveness	0.40	Management mechanism	0.25	8.0-9.0	3.0-5.0	7.0-8.0	6.0-7.0
		Assessment mechanism	0.20	3.0-5.0	8.0-9.0	7.0-8.0	6.0-7.0
		Learning mechanism	0.20	7.0-8.0	6.0-7.0	3.0-5.0	8.0-9.0
		Competition mechanism	0.15	6.0-7.0	7.0-8.0	8.0-9.0	3.0-5.0
		Excitation mechanism	0.20	3.0-5.0	8.0-9.0	6.0-7.0	7.0-8.0

Table 3 presents grey correlation analysis statistics of enterprise human resource competitiveness obtained based on abovementioned algorithm and model.

TABLE 3 Grey correlation analysis values of enterprise human resource competitiveness

	Rank			
	Fairly good	Good	Ordinary	Bad
Positive grey correlation	0.785	0.817	0.633	0.432
Negative grey correlation	0.713	0.625	0.804	0.898
Comprehensive grey correlation	0.548	0.631	0.383	0.188

According to the analysis procedure and result, the human resource competitiveness of the enterprise is in good condition.

4 Conclusion

Evaluation of enterprise human resource competitiveness

was a complex problem of system decision analysis with multiple levels, attributes and objects influenced by multiple factors. This paper put forward an improved evaluation index system of enterprise human resource competitiveness via analysis of relative restrictive factors in the process of implementation plan of enterprise human resource. Then standardize different categories of evaluation indexes under the evaluation index system and put forward an improved grey correlation analysis model of enterprise human resource competitiveness based on the theory of grey system. This model obtains the grades of enterprise human resource competitiveness or the optimal implementation plan in multiple implementation plans. The physical significance of the model is clear and definite, the calculation in the model is easy, and will provide strong support for implementation of computer aided intelligent analysis method of evaluation of enterprise human resource competitiveness and implementation of the system. In the paper, the model and algorithm proved feasible via case analysis.

References

[1] Jill C, Ulrich D 1996 Human resource roles Creating value, not rhetoric *Human Resource Planning* **19**(3) 38-49
[2] Ulrich D, Beatty D 2001 From partners to players: extending the HR playing field, *Human Resource Management* **40**(4) 293-307
[3] Ulrich D, Brockbank W, Yeung A K, GLake D 2006 Human resource competencies: An empirical assessment *Human Resource Management* **34**(4) 473-95
[4] Liu S C, Ge L R 2008 Study on human resource performance management system based on competency model *Economist* (6)
[5] Lawler E E, Boudreau J W2009 What makes HR a strategic partner? *People & Strategy* **32**(1) 14-22
[6] Cui X, Qu J J, Zhang Y M 2013 Research review and path analysis

of diversified influence of human resource *Science of Science and Management of S & T* **34**(9) 172-80
[7] Liang Q Z, Liu Z, Wu Y 2008 Multiple categories, measurement and major research method *Management Review* (11) 51-6
[8] Richard O C 2000 Racial diversity, business strategy and firm performance: A resource-based view *The Academy of Management Journal* **43**(2) 164-77
[9] Olson B J, Parayitam S, Bao Y J 2007 Strategic decision making-the effects of cognitive diversity, conflict and trust on decision outcomes *Journal of Management* **33**(2) 196-222
[10] Lawler E E, Boudreau J W 2006 HR support for corporate boards *Human Resource Planning* **29**(1) 15-24
[11] Rao C, Xiao X, Peng J 2007 Novel combinatorial algorithm for the

problems of fuzzy grey multi-attribute group decision making *Journal of Systems Engineering and Electronics* **18**(4) 774-80

[12] Wang T, Yang A, Bui L 2013 Mechanism scheme design based on multi-attribute extension gray relevant optimized decision-making model *Systems Engineering - Theory & Practice* **33**(9) 2321-9

[13] Deng J L 2002 Estimate and decision of grey system *Wuhan: Huazhong University of Science&Technology Press*

[14] Zhang J J, Wu D S, Olson D L. 2005 The method of grey related analysis to multiple attribute decision making problems with interval numbers *Mathematical and Computer Modelling* **42** 991-8

Authors

Xiuli Li, born on December 17, 1977, Funing, Hebei, China

Current position, grades: director of teaching division, instructor.
University studies: insurance.
Scientific interest: insurance, management.
Publications : 19papers.
Experience: 14 years teacher in Hebei Normal University of Science & Technology

Zhang Yuhong, born on February 9, 1970, Hebei, China

Current position, grades: associate professor of College of Education, Hebei Normal University of Science & Technology.
University studies: B.Sc. of education in Hebei Normal University of Science & Technology, M. Sc. of education in Hebei University.
Scientific interest: education technology, vocational education, and information education.
Publications : Zhang Yuhong, Gu Liwei and Wang Yubin. Research of Proxy Cache Algorithm in Multi-media Education System. International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.2 (2014), pp.15-26 http://dx.doi.org/10.14257/ijmue.2014.9.2.02

Sujuan Zhao, born on February 20, 1985, Liaocheng, Shangdong, China

Current position: assistant.
University studies: master's degree in electronic commerce.
Scientific interest: big data, cloud computing.
Experience: 3 years teaching experience.

A membership degree algorithm of collaborative design roles in distributed design transaction

Qiaojuan Feng*, Yongjiao Wang

Henan University of Urban Construction, LongXiang Road, Pingdingshan City, Henan Province, China

Abstract

As there is a priority in multiple design transaction and multiple design roles in the distributed design traction, this paper studies the distribution hierarchy of collaborative design roles and proposes a membership degree algorithm based on grey related analysis. In this algorithm, design constraint degree about the design roles is acquired through grey related analysis. Under different design constraint degree, design roles and incidence coefficient are acquired together with corresponding matrix. Then the membership degree can be available by the constraint degree and the matrix. Empirical test proves the efficacy and practicability of this algorithm.

Keywords: distributed design, collaborative design, membership degree, algorithm, artificial intelligence

1 Introduction

With the rapid development of computer science, distributed collaborative design has become a big concern of intelligent design. The quick design of huge and complicated modern manufacturing equipment requires multiple design roles in the process of distributed design or manufacturing, as these designs need collaboration of a team or the design is distributed in different enterprises or regions [1-3].

Some researchers have already studied the distributed design and multiple roles collaborative design with satisfactory progress. But as the weight and hierarchy of the design is limited, there are few researches involving fuzzy design roles. Design constraints and standards need to be considered for the hierarchy of fuzzy design roles. Therefore, it is significant to determine the priority of different roles in the collaborative design, which will ease the coupling and conflicts in multiple design and support complicated design that promote the development of artificial intelligent design.

Currently, comprehensive evaluation method and Access Hierarchy Process (AHP) [8-10] are major ways to determine the priority of design roles. However, these methods rely much on people's subjective judgment and experience that may overlook constraint information or uncertain information, resulting in a biased design. Therefore, this paper analyses the membership degree of collaborative design roles based on grey system theory [11-14]. In this theory, the more similar the curves of the sequence are to each other, the more incidences they have. Multiple design constraint is set as the standard for membership degree. The incidence coefficient and incidence matrix between design constraint and design roles are acquired. Then this paper proposes a multiple

design membership degree algorithm based on grey incidence analysis and apply it to empirical study to prove its feasibility.

2 Multiple design roles priority distribution standard

Multiple design roles distribution is key to the collaborative design, the process of which involves in weight distribution affected by various factors. Some factors can be measured up accurately in a quantitative way. Some call for qualitative analysis mixed with uncertain design factors. Thus, multiple design roles distribution is a system decision-making problem with multi-attribute.

Multiple design roles distribution based on grey incidence analysis requires the calculation of grey incidence degree of various design roles and the sequence of roles characteristic and judgment standard. A combination of quantitative and qualitative analysis will make it easier to standardized and do statistics of the priority distribution standard and eliminate human factors so that the result will be more scientific, objective and accurate. Grey incidence analysis serves to the nature of collaborative design represented by priority distribution standard of multiple design roles.

2.1 DETERMINATION OF MULTIPLE DESIGN ROLES DISTRIBUTION STANDARD

Factors that affect the priority distribution of design roles are multifaceted, including organization factor, intrapersonal factor, theme factor, object factor, knowledge sharing, intelligent integration, design demand, etc. Some factors are fuzzy and uncertain. What's more, it also looms large to transmit these factors to an information

---

model that is easy to be identified by the computer, which will ease the priority distribution of design roles.

There is a necessity to set standard and construct a model for standard of the priority distribution of roles. Standards are set based on the abstract and fuzzy constraint factors of collaborative design in the way of semantic segmentation. The purpose is to acquire the minimum, complete and independent unit of constraint information of multiple design roles and transmit the semantics by units to standards that are coped easily.

This paper introduces the concept of unit to describe the priority distribution standard. The priority distribution standard O, standard characteristics C and value of O about C consist a three-element set $J = (O,C,V)$, which serves as the unit of priority distribution standard with one dimension. As the priority distribution standard has various characteristics, they are usually multi-dimensional. Here prescribes that the standard O, n characteristics $c_1, c_2, ..., c_n$ and corresponding value $v_1, v_2, ..., v_n$ consist of rows and lines with n dimensions. It is a standard unit with n dimensions:

$$J = (O,C,V) = \begin{bmatrix} O & c_1 & v_1 \\ & c_2 & v_2 \\ & & \\ & c_n & v_3 \end{bmatrix}. \quad (1)$$

To ensure the reliability of the membership degree distribution, authenticity and completeness of priority distribution standard should be guaranteed. After decomposition, what the original priority distribution standards stand for should be passed down to decomposed standards and the constraint information shouldn't be overlapped. It is expressed by:

$$(\forall J_s, J_t) J_s \cap J_t \wedge (J_s \in J, J_t \in J) = \varnothing, \; s \neq t.$$

2.2 MODEL OF PRIORITY DISTRIBUTION STANDARD

Suppose there are M priority distribution standards and none of them are overlapped in information, and there are K decision subjects to do the grey decision analysis. Here gets the grey decision analysis sequence for standard J_i:

$$\Psi_i = \left(\varphi_i(1), \varphi_i(2), ..., \varphi_i(K) \right), \; i = 1, 2, ..., M.$$

Construct the ideal grey incidence sequence Ψ_Θ of priority distribution standard:

$$\Psi_\Theta = \left(\varphi_\Theta(1), \varphi_\Theta(2), ..., \varphi_\Theta(K) \right) = \\ \left(\max_{1 \leq i \leq M} \left(\varphi_i(1) \right), \max_{1 \leq i \leq M} \left(\varphi_i(2) \right), ..., \max_{1 \leq i \leq M} \left(\varphi_i(K) \right) \right), \quad (2)$$

If $\varphi_i(j)$ is a quantitative value, then the grey incidence coefficient between sequence Ψ_i and sequence Ψ_Θ about j is described as:

$$\delta_i(j) = \\ \frac{\min_i \min_j \left| \varphi_\Theta(j) - \varphi_i(j) \right| + \beta \max_i \max_j \left| \varphi_\Theta(j) - \varphi_i(j) \right|}{\left| \varphi_\Theta(j) - \varphi_i(j) \right| + \beta \max_i \max_j \left| \varphi_\Theta(j) - \varphi_i(j) \right|}, \quad (3)$$

where β is the identification coefficient, and $\beta = 0.5$.

If $\varphi_i(j)$ is fuzzy and uncertain value interval, that is when $\varphi_i(j) = \left[\varphi_i^a(j), \varphi_i^b(j) \right]$ and fits $\varphi_i^a(j) \leq \varphi_i^b(j)$, then the grey incidence coefficient between sequence Ψ_i and sequence Ψ_Θ about j is described as:

$$\delta_i(j) = \frac{\min_i \min_j \sqrt{\frac{\left| \varphi_\Theta^a(j) - \varphi_i^a(j) \right|^2 + \left| \varphi_\Theta^b(j) - \varphi_i^b(j) \right|^2}{2}} + \beta \max_i \max_j \sqrt{\frac{\left| \varphi_\Theta^a(j) - \varphi_i^a(j) \right|^2 + \left| \varphi_\Theta^b(j) - \varphi_i^b(j) \right|^2}{2}}}{\sqrt{\frac{\left| \varphi_\Theta^a(j) - \varphi_i^a(j) \right|^2 + \left| \varphi_\Theta^b(j) - \varphi_i^b(j) \right|^2}{2}} + \beta \max_i \max_j \sqrt{\frac{\left| \varphi_\Theta^a(j) - \varphi_i^a(j) \right|^2 + \left| \varphi_\Theta^b(j) - \varphi_i^b(j) \right|^2}{2}}}. \quad (4)$$

The grey incidence degree between sequence Ψ_i and sequence Ψ_Θ are:

$$\lambda(i) = \frac{1}{K} \sum_{j=1}^{K} \delta_i(j). \quad (5)$$

The priority degree of standard J_i is described as:

$$\kappa_i = \lambda(i) / \sum_{i=1}^{M} \lambda(i). \quad (6)$$

The priority degree of all priority distribution standards of multiple design roles are expressed by:

$$\kappa = \left(\kappa_1, \kappa_2, ..., \kappa_M \right)^T. \quad (7)$$

3 Calculation method and algorithm of membership degree of multiple design roles

3.1 CALCULATION METHOD OF MEMBERSHIP DEGREE

Suppose there are H collaborative design roles in the distributed design and K decision subjects to do the grey decision analysis. The grey decision sequence for roles J_t is described as:

$$\Phi_r = \left(\phi_r(1), \phi_r(2), ..., \phi_r(K) \right), \; r = 1, 2, ..., H.$$

In order to measure better the membership degree, here sets the priority distribution standard Ψ_i as the ideal grey

incidence sequence Φ_Θ. Calculate the grey incidence degree between design roles sequence Φ_r and ideal grey incidence sequence Φ_Θ as well as the incidence coefficient $\xi_r(k)$ between priority distribution standard and design roles.

Suppose the ideal grey incidence sequence Φ_Θ is expressed as:

$$\Phi_\Theta = \Psi_i = \left(\varphi_i(1), \varphi_i(2), ..., \varphi_i(K) \right). \tag{8}$$

If $\varphi_i(j)$ is a quantitative value, then the grey incidence coefficient between sequence Φ_r and sequence Ψ_i about j is described as:

$$\tau_i^r(j) =$$

$$\frac{\min\limits_i \min\limits_r \left| \varphi_i(j) - \phi_r(j) \right| + \beta \max\limits_i \max\limits_r \left| \varphi_i(j) - \phi_r(j) \right|}{\left| \varphi_i(j) - \phi_r(j) \right| + \beta \max\limits_i \max\limits_r \left| \varphi_i(j) - \phi_r(j) \right|}. \tag{9}$$

If $\varphi_i(j)$ is fuzzy and uncertain value interval, that is when $\phi_r(j) = \left[\phi_r^a(j), \phi_r^b(j) \right]$ and fits $\phi_r^a(j) \le \phi_r^b(j)$, then the grey incidence coefficient between sequence Φ_r and sequence Ψ_i about j is described as:

$$\tau_i^r(j) = \frac{\min\limits_i \min\limits_r \sqrt{\frac{\left| \varphi_i^a(j) - \phi_r^a(j) \right|^2 + |\varphi_i^b(j) - \phi_r^b(j)|^2}{2}} + \beta \max\limits_i \max\limits_r \sqrt{\frac{\left| \varphi_i^a(j) - \phi_r^a(j) \right|^2 + |\varphi_i^b(j) - \phi_r^b(j)|^2}{2}}}{\sqrt{\frac{\left| \varphi_i^a(j) - \phi_r^a(j) \right|^2 + |\varphi_i^b(j) - \phi_r^b(j)|^2}{2}} + \beta \max\limits_i \max\limits_r \sqrt{\frac{\left| \varphi_i^a(j) - \phi_r^a(j) \right|^2 + |\varphi_i^b(j) - \phi_r^b(j)|^2}{2}}}. \tag{10}$$

The incidence matrix τ between the priority distribution standard and the collaborative design roles:

$$\tau = \begin{bmatrix} \tau_1^1 & \tau_1^2 & \cdots & \tau_1^H \\ \tau_2^1 & \tau_2^2 & \cdots & \tau_2^H \\ \vdots & \vdots & \vdots & \vdots \\ \tau_M^1 & \tau_M^2 & \cdots & \tau_M^H \end{bmatrix}_{MxH}. \tag{11}$$

The absolute grey membership degree sequence of collaborative design roles is:

$$\omega = \kappa^T * \tau = \left(\omega_1, \omega_2, ..., \omega_P \right)^T, \tag{12}$$

The absolute grey membership degree ω_r of the collaborative design roles r is:

$$\omega_r = \sum_{i=1}^M \left(w_i * \tau_i^r \right), \ 1 \le r \le H. \tag{13}$$

The absolute grey membership degree ϖ_r of the collaborative design roles r is:

$$\varpi_r = \omega_r / \sum_{r=1}^H \omega_r. \tag{14}$$

The relative grey membership degree sequence of collaborative design roles is:

$$\varpi = \left(\varpi_1, \varpi_2, ..., \varpi_H \right)^T. \tag{15}$$

3.2 DISTRIBUTION ALGORITHM FOR MULTIPLE MEMBERSHIP DEGREE OF DESIGN ROLES BASED ON GREY INCIDENCE ANALYSIS

As is discussed above, the grey incidence sequence is constructed based on the priority distribution standard of collaborative design roles in the distributed design transaction. Then, the priority distribution standard sequence is set as the ideal grey incidence analysis sequence and absolute membership degree and relative membership degree are available by the improved grey incidence calculation method proposed in this paper. The membership degree distribution algorithm based on grey incidence analysis is described as follows:

Step 1 Construct roles model according to design demand to confirm the design roles and corresponding roles characteristics;

Step 2 Determine the priority distribution standard of collaborative design roles in distributed design transaction based on Chapter 1.1;

Step 3 According to section 1.2, conduct the grey decision analysis abased on priority distribution standard of decision subjects to multiple design roles and get the grey decision sequence;

Step 4 Construct the ideal grey incidence analysis incidence of the priority distribution standard based on Equation (2);

Step 5 According to specific characteristics of standard, get the grey incidence coefficient of priority distribution standard based on Equations (3) and (4);

Step 6 Get the grey incidence degree and priority degree of priority distribution standard based on Equations (5)-(7);

Step 7 Construct the priority distribution standard of multiple design roles as the ideal grey incidence sequence according to section 2.1 based on Equation (8);

Step 8 Get the grey incidence analysis coefficient based on Equations (9-11) and construct corresponding coefficient matrix;

Step 9 Get the absolute membership degree and relative membership degree based on Equations (12)-(15) and determine the hierarchy of roles;

Step 10 Complete the membership degree distribution.

Operation Research and Decision Making

Feng Qiaojuan, Wang Yongjiao

4 Empirical test

This paper tests the algorithm by a supporting system of a product's distributed design. This distributed design system has five design groups as design roles to coordinate tasks. Organization factor, theme factor, object factor, knowledge sharing, intelligent integration, design demand are set as distribution standard for membership degree. The original sequences by grey incidence analysis of priority distribution standard of multiple design roles and collaborative design roles in the distributed design transaction are shown in Tables 1 and 2.

TABLE 1 Grey incidence analysis of priority distribution standard of multiple design roles

Distribution standard	Grey incidence analysis data
Organization factor	9, (7.5,8.5),9,9, (6.5,7.5),9,9, (7.5,8.5)
Theme factor	9,9,9,9,9, (7.5,8.5),9,9
Object factor	(7.5,8.5),9,9,9, (7.5,8.5),9, (7.5,8.5),9
Knowledge sharing	(7.5,8.5),9, (6.5,7.5),9, (6.5,7.5), (7.5,8.5),9, (7.5,8.5)
Intelligent integration	(7.5,8.5), (5.5,6.5), (6.5,7.5), (7.5,8.5),9, (7.5,8.5), (6.5,7.5), (7.5,8.5)
Design demand	9, (6.5,7.5), (7.5,8.5), (7.5,8.5), (7.5,8.5),9, (6.5,7.5), (7.5,8.5)

TABLE 2 Grey incidence analysis of collaborative design roles

Collaborative design roles	Grey incidence analysis data
Design roles 1	(5.5,6.5), (5.5,6.5), (6.5,6.5), (6.5,7.5), (7.5,8.5), (6.5,7.5),9, (6.5,7.5)
Design roles 2	(6.5,7.5), (5.5,6.5), (7.5,8.5), (5.5,6.5), (6.5,7.5), (7.5,8.5), (6.5,7.5), (5.5,6.5)
Design roles 3	9, (7.5,8.5), (7.5,8.5),9,9,9, (6.5,7.5), (7.5,8.5)
Design roles 4	(5.5,6.5), (5.5,6.5), (4.5,5.5), (5.5,6.5), (6.5,7.5), (5.5,6.5), (6.5,7.5), (5.5,6.5)
Design roles 5	(6.5,7.5), (5.5,6.5), (5.5,6.5), (5.5,6.5), (7.5,8.5), (5.5,6.5), (6.5,7.5), (5.5,6.5)

According to the description of algorithm in section 2.2, the grey incidence degree and priority degree are available and shown in Table 3.

TABLE 3 The grey incidence degree and priority degree of distribution standard

Distribution standard	Calculated data for grey incidence analysis
Grey incidence degree	0.926, 0.961,0.942,0.780, 0.731,0.794
Grey priority degree	0.180,0.187,0.183,0.152, 0.142,0.155

Select several distribution standards as ideal grey incidence sequence. According to the algorithm described in section 2.2, absolute membership degree and relative membership degree are available and shown in Table 4.

TABLE 4 Membership degree by grey incidence analysis

Membership degree	Calculated data for grey incidence analysis
Absolute membership degree	0.542, 0.581,0.812,0.458, 0.416
Relative membership degree	0.193,0.207,0.289,0.163, 0.148

From Table 4, it is clear that design roles 3 has the highest membership degree, design roles 2 and 1 have the medium membership degree and design roles 4 and 5 have the lowest membership degree. Therefore, the design roles with a higher membership degree should enjoy a higher priority so that it can support the distributed collaborative design better.

5 Conclusion

This paper proposes a membership degree algorithm based on grey related analysis. The priority of the multiple design roles is the standard for the judgment of roles distribution. This paper takes into consideration various factors that affect the distributed design transaction and enables the membership degree to meet the demand of the coordinative design to provide support for the distributed design. Based on grey related analysis, this paper improves the algorithm and makes it more scientific, objective and accurate. Finally, empirical test proves the efficacy and practicability of this algorithm.

References

[1] Chen C T 2001 A fuzzy approach to select the location of the distribution center *Fuzzy Sets and System* **18**(1) 65-73

[2] Zhou A, Li W 2013 Maturity degree evaluation method of components in plane collaborative design *China Mechanical Engineering* **24**(1) 61-5

[3] Wu J, Qiu Q, Feng P, Gao F 2005 Management strategy of product data in distributed collaborative design environment *Journal of Zhejiang University (Engineering Science)* **39**(10) 1465-9 *(in Chinese)*

[4] Sheu R, Chang Y, Yuan S 2001 Managing and sharing collaborative files through WWW *Future Generation Computer Systems* **17** 1039-49

[5] Zou L, Guo D, Gao H, Sun C 2010 Fuzzy Evaluation Method of Design Maturity for Collaborative Product Development *Journal of Computer-aided Design & Computer Graphics* 2010 **22**(5) 791-6

[6] Dix A, Rodden T 1997 Modelling Versions in Collaborative Work *IEEE Proceedings Software Engineering* **114**(4) 195-205

[7] Wu D, Wang X, Wei Z, Li H 2002 Distributed product data management on collaborative service platforms *Journal of Tsinghua University: Natural Science* **42**(6) 791-4 *(in Chinese)*

[8] Durán O, Aguilo J 2008 Computer-aided machine-tool selection based on a Fuzzy-AHP approach *Expert Systems with Applications* **34**(3) 1787-94

[9] Hou C, Liu H 2005 The application of fuzzy multiple-attribute decision making model in water power station engineering *Journal of Wuhan University (Engineering and Science)* **38**(5) 25-7 *(in Chinese)*

[10] Zeng G, Jiang R, Huang G, Xu M, Li J 2007 Optimization of wastewater treatment alternative selection by hierarchy grey relational analysis *Journal of Environmental Management* **82**(2) 250-9

[11] Karmakar S, Mujumdar P P 2006 Grey fuzzy optimization model for water quality management of a river system *Advances in Water Resources* **29**(7) 1088-1105

[12] Fan Z P, Ma J, Zhang Q 2002 An approach to multiple attribute decision making based on fuzzy preference information on alternatives *Fuzzy Sets and Systems* **131**(1) 101-6

[13] Zhang J J, Wu D S, Olson D L 2005 The method of grey related analysis to multiple attribute decision making problems with interval numbers *Mathematical and Computer Modelling* **42** 991-8

[14] [Li L, Guo Q, Li Y 2007 Quality function deployment based on grey relational analysis *Computer Integrated Manufacturing Systems* **13**(12) 2469-72

Authors

Qiaojuan Feng, born in December, 1977, Pingdingshan, Henan, China

Current position, grades: experimentalist, teacher at Henan University of Urban Construction.
University studies: Master's degree in engineering in computer science in Huazhong University of Science and Technology.
Scientific interest: computer technology and computer networks.
Publications number or main: 3.
Experience: Teacher at Henan University of Urban Construction since 2001.

Yongjiao Wang, born in February, 1977, Xinxiang, Henan, China

Current position, grades: associate professor, teacher at Henan University of Urban Construction.
University studies: M.S. degree in computer science in ZheJiang University.
Scientific interest: image processing and computer applications.
Publications: 2.
Experience: Teacher at Henan University of Urban Construction since 1999.

Operation Research and Decision Making

Actor-Critic reinforcement learning based on prior knowledge

Zhenyu Yang*

Qilu University of Technology Jinan, China

Received 1 March 2014, www.cmnt.lv

Abstract

In order to improve the incremental learning algorithm Actor-Critic learning efficiency, from a policy learning, introduce experience sample data into incremental Actor-Critic algorithm, make effective use of the useful information contained in the sample data of experience in the learning process. Given the recursive least-squares temporal difference, RLSTD (λ) algorithm and incremental least-squares temporal difference, iLSTD (λ) algorithms are able to make good use sample data collected in the past, respectively RLSTD and iLSTD algorithm is applied to policy evaluation Critic's. Then, Critic learned value function based on RLSTD or iLSTD algorithm, Actor gradient update strategy based on conventional parameters, so the improvement of Critic effectiveness assessment will help Actor to improve strategy-learning performance. Finally, simulation studies on two control problems with continuous state space, analyse the impact of different parameters on the performance of the learning algorithm and verify its effectiveness.

Keywords: actor-critic, RLSTD, on-policy, ILSTD

1 Introduction

Reinforcement learning as an effective method for solving a class of Markov decision model problem, has wide application in optimization and control. As a kind of reinforcement learning in the field of research, policy gradient reinforcement learning is a way to search directly on the strategy parameter space, this method overcomes the reinforcement learning algorithm based on the value of the function can not guarantee the convergence of shortcomings. However, due to the variance in the gradient estimation process is too large, resulting in policy gradient reinforcement learning method converges too slowly, thus impeding the policy gradient reinforcement learning method is widely used. By combining the value function method, come up with an Actor-Critic (AC) reinforcement learning method, AC approach combines the advantages of fast learning based RL value function method and strategies gradient RL is easy to converge. You can reduce the gradient estimation variance is an important strategy gradient reinforcement learning method has wide application in solving large-scale and high-dimensional Markov Decision learning control problems.

Actor-Critic learning structure shown in Figure 1. Actor-Critic learning consists of policy evaluation and policy improvement, which, Critic (evaluator) present a prediction problem, according to the time difference learning to estimate the value of the function, Actor (actuators) presents a control problem, update the policy parameters dynamically according to learn the value of the function. But the policy iteration method is different, Actor-Critic gradient method to update the policy parameters in the direction of increasing the expected returns, but policy iteration is by maximizing the value of the function on the action space to update the policy.
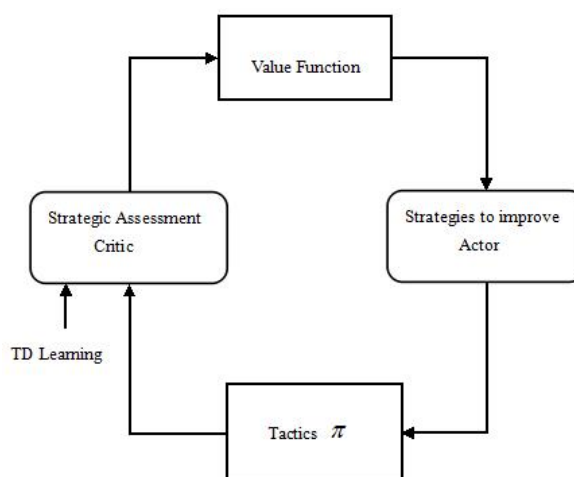
FIGURE 1 Actor-Critic architecture

Actor-Critic is an online on-policy learning algorithm, at each time step agent collects samples in accordance with current policy, based on the sample, Critic use the time difference algorithm to learn value function, Actor update strategy based on the estimated value function. When the policy updates, agent on the basis of new strategy to re-collect samples for learning, and discard the old sample, in order to obtain a satisfactory strategy, agent requires constant sampling more data, so that useful information is not contained in the old sample is fully utilized, not only waste a lot of sample data, and the impact of the algorithm's learning speed and low learning efficiency. One kind of AC algorithm can effectively improve the learning efficiency is experience playback, which can take full advantage of the useful information contained in the

---

* *Corresponding author* e-mail: yang_zhenyu@163.com

sample data experience. Therefore, this paper will examine how to introduce experience in the sample data in the AC learning process to improve the learning performance of the algorithm.

2 Efficient data use in incremental Actor-Critic algorithms

A major part of the Actor-Critic learning algorithm is policy evaluation, based on sample data online estimated strategy corresponding value function. TD learning to update the estimates based on the estimated value already exists, is a Bootstrapping method. AC in incremental learning algorithm, at each time step Critic learning value function according to the TD, Actor according to a biased conventional gradient to update policy parameters, in order to facilitate comparison with the algorithms behind this algorithm is denoted as AC-TD.

In the policy evaluation, TD learning at each time step only based on the current sample to estimate value function, a small amount of calculation, but it cannot effectively use the last sample data, you need a lot of time step to learn an accurate estimate. In order to improve data availability, while overcoming the difficulties of learning step design. Bradtke put forward a linear value function approximation based on LSTD learning algorithm, the algorithm directly to Markov decision process value function approximation of mean square error as the performance index [4]. In order to further improve the generalization performance of LSTD algorithm, Boyan et. combined eligibility trace λ extended for LSTD and proposed LSTD (λ) algorithm [5].When using the function approximator solving a value function parameters, although LSTD (or LSTD (λ)) algorithm can make more effective use of sample data considerably more than the last TD algorithm to get the correct estimate, however, when asked questions has great feature space and the time required for online estimation, LSTD algorithm at each time step requires a lot of computing, resulting in decreased learning efficiency. Therefore, LSTD online learning algorithm for solving problems with a large number of feature vectors is impractical. In order to reduce the computational burden LSTD (λ) algorithm, Xu et proposed a Recursive LSTD (λ) - RLSTD (λ) algorithm [9]. RLSTD (λ) algorithm is more suitable for online learning, and more effective use sample data than TD algorithm. In addition, Geramifard et put forward an Incremental LSTD, iLSTD learning algorithm [6] and then combined eligibility trace to extend iLSTD and proposed iLSTD (λ) algorithm [7]. iLSTD (λ) algorithm played a role of a compromise between TD algorithm in a low utilization rate of the data and a large computation complexity of LSTD algorithm, it is more effective to use sample data than TD to obtain a good approximation, and compared with LSTD algorithm, reducing the computational complexity. In summary, in order to overcome the problem of low utilization of data caused by TD Critic assessment in incremental algorithm in AC, as

well as improving the efficiency of learning algorithms, in this paper, starting at learning strategies, the introduction of experience in policy evaluation sample incremental algorithm in AC. The two variants LSTD (λ) method (i.e. RLSTD (λ) and iLSTD (λ)) were applied to the Critic proposes two new incremental AC algorithm, namely AC-RLSTD and AC-iLSTD. Critic use RLSTD (λ) or iLSTD (λ) algorithm to estimate the value function, Actor update strategy based on a conventional gradient obtained by the TD error. Critic use two variants LSTD (λ) method in policy evaluation to improve the performance Critic learning assessments, and thus can more efficiently update Actor strategy parameters based on the assessment results.

3 Based on TD incremental Actor-Critic learning

Policy gradient reinforcement learning goal is to learn an optimal or suboptimal strategy, namely to estimate expected returns relative to the gradient of the policy parameters, and then use the gradient parameter adjustment the policy parameters. According to the policy gradient theorem, the expected return on the policy parameters 'Vanilla' gradient can be expressed as [14]:

$$\frac{\partial J(\theta)}{\partial \theta} = \sum_s d^\pi(s) \sum_a \frac{\partial \pi(a|s,\theta)}{\partial \theta} Q^\pi(s,a) =$$
$$\sum_s d^\pi(s) \sum_a \nabla_\theta \pi(a|s,\theta)(Q^\pi(s,a) - b^\pi(s))$$, (1)

where $b^\pi(s)$ represents a baseline, other, Bhatnagar et al [15,9] and Sutton et al [16] proved $Q^\pi(s,a) - b^\pi(s)$ that you can use a compatible function approximation $f_{\bar{w}}(s,a) = \bar{w}^T \psi_{sa}$ to express, where, $\bar{w}$ is a parameter vector, when $b^\pi(s) = V^\pi(s)$,The device is compatible function approximation of minimum mean square error [15,9]. Therefore, the Equation (1) can be written as:

$$\nabla_\theta J(\theta) = \frac{\partial J(\theta)}{\partial \theta} =$$
$$\sum_s d^\pi(s) \sum_a \frac{\partial \pi(a|s,\theta)}{\partial \theta}(\bar{w}^{*T} \psi_{sa}) =$$
$$\sum_s d^\pi(s) \sum_a \pi(a|s,\theta) \psi_{sa} A^\pi(s,a)$$

$$\psi_{sa} = \nabla_\theta \log \pi(a|s,\theta) ,$$ (2)

where, $A^\pi(s,a) = Q^\pi(s,a) - V^\pi(s)$ is an advantage function.

From the above Equation (2) can be seen, $\nabla_\theta J(\theta)$ calculation depends on the advantages of function $A^\pi(s,a)$, however, in [14] pointed out. Advantage of $A^\pi(s,a)$ function not just in the approximation $f_{\bar{w}}^\pi(s,a)$ is obtained based on TD learning. To this end, Bhatnagar et al [15, 9] and Morimura [11] by TD error status value function to construct the advantages of function.

According to Bellman equation, TD error is defined as: $\delta_t = r_{t+1} + \gamma V^\pi(s_{t+1}) - V^\pi(s_t)$. Text [15, 9, 11] proved that δ_t consistent estimate of the advantages of function $A^\pi(s, a)$.

However, calculate δ_t need to a known state value function, now use a linear function approximation to estimate the value function, the estimated state value function is $\hat{V}^\pi(s) = \varphi(s)^T \omega$. TD error can be estimated according to $\hat{\delta}_t = r_{t+1} + \gamma\varphi(s_{t+1})^T \omega_t - \varphi(s_t)^T \omega_t$, where ω_t represents the parameter vector at t times update. In incremental AC algorithm, Critic on the basis of TD error estimated to update value function parameters, Actor biased according to a conventional gradient update strategy parameters:

$$\omega_{t+1} = \omega_t + \alpha_t \hat{\delta}_t \varphi(s_t) \,, \tag{3}$$

$$\theta_{t+1} = \theta_t + \beta_t \hat{\delta}_t \psi_{s_t a_t} \,, \tag{4}$$

where, α_t and β_t are the Critic and Actor update step, meet the following conditions:

$$\sum_t \alpha_t = \sum_t \beta_t = \infty$$
$$\sum_t \alpha_t^2, \sum_t \beta_t^2 < \infty \,. \tag{5}$$
$$\beta_t = o(\alpha_t)$$

From the Equation (3) can be seen, update the parameter ω is only determined by the data that current observed in the sample, after the current update is completed, the sample is discarded. Consider the case of non-sparse feature representation, Critic at each time step only needs to calculate the amount of $O(k)$, but, Critic wasted a lot of samples, resulting in the learning process needs to continue to collect more data in order to obtain a satisfactory sample gradient estimates.

4 Based RLSTD (λ) or ILSTD (λ) incremental Actor-Critic learning

TD playback experience is a way to improve the effectiveness of learning data [5]. For example, at each time step, LSTD (λ) method put all observations TD update set to zero to solve the value function parameters. However, the data validity LSTD (λ) method at the expense of a large amount of computation. In order to balance the relationship between data validation and calculation effectiveness of the Critic assessment, respectively RLSTD (λ) and iLSTD (λ) is introduced into Critic and proposes two new incremental AC algorithm.AC algorithm proposed in, Critic according RLSTD (λ) or iLSTD (λ) algorithm to learn the value

function, Actor estimate $\hat{\nabla}_\theta J(\theta)$ parameter update strategy based on a regular gradient.

4.1 BASED RLSTD (λ) TO ASSESS THE CRITIC

Let $u_t(\omega)$ be the sum of the TD updated within t times, according to TD and the definition of the error function of the state value, are:

$$u_t(\omega) = \sum_{n=1}^t u_n(\omega) = \sum_{n=1}^t z_n (r_{n+1} + \gamma\varphi(s_{n+1})^T \omega - \varphi(s_n)^T \omega)$$
$$= \sum_{n=1}^t z_n r_{n+1} - \sum_{n=1}^t z_n (\varphi(s_n) - \gamma\varphi(s_{n+1}))^T \omega = b_t - \tilde{A}_t \omega \tag{6}$$

where, $z_n = \gamma\lambda z_{n-1} + \varphi(s_n)$ is a qualified track, $\lambda \in [0, 1]$ is a parameter set in advance. Online learning, LSTD observed data into each vector and matrix b, $\tilde{A}$, then order sum DT update to calculation parameters, and order the Equation (6) is equal to 0, to obtain a new parameter: $\omega_{t+1} = \tilde{A}_t^{-1} b_t$. A and b when the update is completed, observation of the sample data is ignored, and at this time there is no loss of information.

When b and $\tilde{A}$ of the update is completed, observation of the sample data is ignored, and at this time there is no loss of information. In LSTD algorithm, each time step in the computational complexity of the inverse matrix is A, large computational burden. To this end, the policy evaluation incremental AC algorithms in this section use RLSTD (λ) algorithm to solve LSTD (λ) the computational complexity of the problem.

Order $F_t = A_t^{-1}$, $F_0 = \rho I$, $G_{t+1} = F_{t+1} z_t$, wherein A is a positive number, is a unit matrix. Find the matrix inverse theorem [12] shows that the value of the function parameter update rules is as follows:

$$G_{t+1} = F_t z_t / (1 + (\varphi(s_t)^T - \gamma\varphi(s_{t+1})^T) F_t z_t) \,, \tag{7}$$

$$\omega_{t+1} = \omega_t + G_{t+1}(r_t - (\varphi(s_t)^T - \gamma\varphi(s_{t+1})^T)\omega_t) \,, \tag{8}$$

$$F_{t+1} = F_t - \frac{F_t z_t (\varphi(s_t)^T - \gamma\varphi(s_{t+1})^T) F_t}{1 + (\varphi(s_t)^T - \gamma\varphi(s_{t+1})^T) F_t z_t} \,. \tag{9}$$

Calculation of each time step is $O(k^2)$. There is an additional parameter in Critic evaluation, the initial value ρ of the initial covariance matrix F_0. As Xu et al [9] above, the initial constant ρ plays an important role in convergence RLSTD (λ) of the algorithm. Performance RLSTD (λ) has a larger initial constant of ρ is similar to the LSTD algorithm. In some cases, RLSTD having a smaller value of ρ RLSTD algorithm is faster than having a larger value of ρ algorithm convergence speed, can refer to this phenomenon [13] Theoretical analysis. In [13] Moustakides proposed a recursive least squares method, noting that at moderate or high SNR should use a relatively

small initial matrix, and in the low SNR choose a relatively large initial matrix for optimal performance. Based on the above analysis, similar to RLSTD (λ) algorithm, this paper will analyse the initial constant value ρ in the following simulation impact AC algorithm performance based RLSTD (λ) incremental.

4.2 ILSDR BASED (λ) TO ASSESS THE CRITIC

Unlike LSTD algorithms, iLSTD (λ) algorithm to solve the value function parameters in incremental form.TD update error until all observations data reduced to zero. In iLSTD (λ) algorithm, the update with the status transition and value function parameters occurred, $u_t(\omega_t)$ calculation of the incremental update. Further, based on the observed state transition and returns a new, $\tilde{A}_t$ and b_t are incrementally updated:

$$b_t = b_{t-1} + r_{t+1}z_t = b_{t-1} + \Delta b_t,$$ (10)

$$\tilde{A}_t = \tilde{A}_{t-1} + z_t(\varphi(s_t) - \gamma\varphi(s_{t+1}))^T = \tilde{A}_{t-1} + \Delta\tilde{A}_t,$$ (11)

where, Δ represents the amount of change in adjacent time variable. Given $\tilde{A}_t$ and b_t, there is:

$$u_t(\omega_t) = u_{t-1}(\omega_t) + \Delta b_t - (\Delta\tilde{A}_t)\omega_t.$$ (12)

LSTD and TD algorithm updates the parameter vector ω is composed of all the elements, and iLSTD (λ) algorithm updates only ω of all the constituent elements of a small portion. For example, consider updating the first A elements:

$$\omega_{t+1} = \omega_t + \alpha_t u_t(i)e_i,$$ (13)

where, $u_t(i)$ is the first element of u_t, i, e_i is a column vector, e_i mere element in row i is 1, the remaining behaviour 0.Then incremental form A is:

$$u_t(\omega_{t+1}) = u_t(\omega_t) - \alpha_t u_t(i)\tilde{A}_t e_i.$$ (14)

According to Equation (13) and (14),each time you select an element update, repeated many times, you can complete the updated parameter vector all elements of ω .Thus, an element of choice here will inevitably encounter problems, choose which element to update it? Geramifard and put forward two commonly used feature selection mechanism in [7]: random selection mechanism and greedy selection mechanism. Random selection mechanism based iLSTD (λ) (iLSTD-random) algorithm can converge to a TD (λ) the same result, and the selection mechanism based iLSTD greedy (λ) (iLSTD-greedy) algorithm does not satisfy the convergence because all full conditions, there is no guarantee of convergence. However, Geramifard etc. found in [7] in the performance iLSTD-greedy algorithm is slightly better than iLSTD-random algorithm. In the simulation study the following

learning control problems, we also found that although iLSTD-greedy algorithm converges in theory, the lack of guarantee, however, learning performance selection mechanism based on greedy algorithm is superior to AC-iLSTD based on random selection mechanism AC-iLSTD algorithm.

The main idea of greedy selection mechanism is to select the elements with the largest sum TD update, i.e. $i = \arg\max(|u_t(i)|)$. Suppose that $\tilde{k}$ represents the number of elements need to be updated at each time step, the iLSTD (λ) is calculated for each time step in the algorithm complexity of $O(k^2 + \tilde{k}k)$. For online learning, when the number represented by the characteristic dimensions are very large, the calculation efficiency iLSTD (λ) is higher than LSTD algorithm, and can be used more effectively than in the past sample data TD algorithm.

4.3 ALGORITHM STEPS

Based on the above analysis, are given based on RLSTD (λ) or iLSTD (λ) AC incremental learning algorithm as follows:

Step 1. Given a random parameters strategy $\pi(a \mid s,\theta)$, feature vector valued functions $\varphi(s)$, there are $\psi_{sa} = \nabla_\theta \log\pi(s,a)$.

Step 2. Initialization vector policy parameters $\theta = \theta_0$, valued functions parameter vector $\omega = \omega_0$, two learning step $\alpha = \alpha_0$ and $\beta = \beta_0$, the discount factor γ , eligibility trace parameter λ , the initial variance matrix parameters ρ ,convergence error ε . Select an initial state $s_0 \in S$, order $z_{-1} = 0$, $\tilde{A}_{-1} = 0$, $u_{-1} = 0$.

Step 3. Select the action according to A, observe the next state B, and get immediate return r_{t+1} .

Step 4. If you are an absorbing state, then s_{t+1} is set to an initial state, so that eligibility trace z_t is a zero vector. Otherwise, update qualifications trace z_t .

Step 5. Critic use RLSTD (λ) algorithm according to the Equation (7) - (9) to update the value of the function parameters ω_t , or the use of iLSTD (λ) algorithm according to the Equations (10) - (14) update value of the function parameter ω_t .

Step 6. Update TD error δ_t value function according to learn, Actor according to Equation (4) to update the policy parameters θ_t .

Step 7. If you terminate the algorithm, otherwise go to Step 3.

5 Simulation study

In this paper inverted pendulum on two proposed algorithms to assess the incremental AC, the proposed AC-

RLSTD (λ) and AC-iLSTD (λ) two algorithms with AC-TD algorithm were compared. For AC-iLSTD algorithm, characteristics were tested using a random selection mechanism AC-iLSTD (AC-iLSTD-random) feature selection algorithm and greedy mechanism AC-iLSTD (AC-iLSTD-greedy) performance of the algorithm, in both feature selection mechanism, the number of elements selected for each update is $\tilde{k} = 1$.

In order to compare different incremental AC algorithm performance, random selection of parametric Gibbs distribution strategy

$$\pi(a|s, \theta) = \frac{e^{\theta^{\mathrm{T}} \varphi_{sa}}}{\sum_{a' \in A} e^{\theta^{\mathrm{T}} \varphi_{sa'}}},$$

where φ_{sa} is a $k \times |A|$ dimensional state - action feature vector. The initial policy parameters θ_0 and ω_0 the initial state value function parameter is set to 0. Critic and Actor step length, respectively [15]:

$$\beta_t = \frac{\beta_0 \cdot \beta_c}{\beta_c + t}, \ \alpha_t = \frac{\alpha_0 \cdot \alpha_c}{\alpha_c + t^{2/3}}.$$

Which, according to the empirical method, $\alpha_0 = 0.1$, $\beta_0 = 0.01$, $\alpha_c = 1$, $\beta_c = 1,000,000$, and step length rule satisfies (1-5), $\gamma = 0.95$, $\varepsilon = 10^{-6}$.

5.1 INVERTED PENDULUM

Inverted pendulum control is a nonlinear, complex system instability control problems in automatic control, artificial intelligence and machine learning methods are usually used to test the performance of different learning. Reinforcement learning environment as a method for solving a class of unknown model complex problems, often used to solve the inverted pendulum control problem, in order to test the performance of reinforcement learning.

This article uses an inverted pendulum [8] described the model, shown in Figure 2. Inverted pendulum problem is solved by applying a force to the car to balance an unknown quality and length of the pendulum, the pendulum system in a small car. Continuous state space, φ vertical angle and angular velocity of the pendulum composition $\dot{\varphi}$: $s = (\varphi, \dot{\varphi})^{\mathrm{T}}$. Action space is discrete from the left edge -50 N, and the right power +50 N 0 N composed of three actions. Simulation, the inverted pendulum system is described by the following equation:

$$\ddot{\varphi} = \frac{g \sin \varphi (m_c + m_p) - (a + m_p l_p \dot{\varphi}^2 \sin \varphi) \cos \gamma}{\frac{4}{3} l_p (m_c + m_p) - m_p l_p \cos^2 \varphi}, \quad (15)$$

wherein, $g = 9.8 \text{ m/s}^2$ is the gravity constant, $m_p = 2.0 \text{ kg}$ is the mass of the pendulum, $m_c = 8.0 \text{ kg}$ is

the mass of the trolley, $l_p = 0.5 \text{ m}$ is the length of the pendulum. Simulation process, using fourth-order Runge - Kutta method to simulate dynamic systems, simulation time step is set 0.1s. When $|\varphi| \le \pi / 2$ smart body was immediately return 0.The period of time from the initial state to a failed state is defined as between a scene, each scene is set to the initial state equilibrium $(0, 0)^{\mathrm{T}} g$ .When $|\varphi| > \pi / 2$ curtain or greater than the length of training 1000, they think the system encountered a failed state, the scene stops, agent get in return is -1.

Simulation, the state of the feature vector consists of a constant and 9 radial basis function consists of:

$$\varphi(s) = (1, \ e^{-\frac{\|s - c_1\|^2}{2\sigma^2}}, \ \cdots, \ e^{-\frac{\|s - c_9\|^2}{2\sigma^2}})^{\mathrm{T}},$$

wherein, $\{c_i\}_{i=1}^9$ is located in the state space of nine points a 3 × 3 two-dimensional lattice on $\left\{-\frac{\pi}{4}, 0, +\frac{\pi}{4}\right\} \times \{-1, 0, +1\}$, $\sigma^2 = 1$. Based learning strategies, according to the expectations of rewards and balanced steps to assess the different incremental AC algorithms, and analyse the impact eligibility trace algorithm parameters and AC-RLSTD initial variance matrix parameters ρ for its performance of the algorithm. In every training to learn the strategy behind an assessment that is based on learning strategies, so that the inverted pendulum from the initial state of the test run was repeated 10 times. When first run time step over 1000, it has been able to successfully control considered inverted pendulum balance. From the same initial policy and state value function parameters, every one AC algorithm test was repeated 20 times.
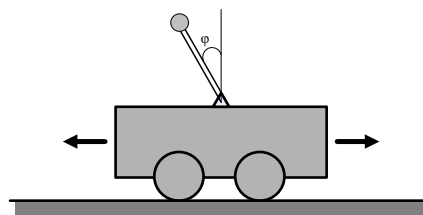
FIGURE 2 Inverted pendulum model

Figure 3 and Figure 4 shows the comparative performance of different AC algorithms under four different values, the horizontal axis represents the number of training curtain ordinate were expected returns and balance the number of steps the average results of 20 runs, where the initial variance matrix parameters AC-RLSTD algorithm ρ to 500. As can be seen from Figures 3 and 4, with the increase of the two AC performance of the algorithm proposed in this section A is significantly improved, the number of steps and the balance obtained by the AC-RLSTD than the AC algorithm step-iLSTD balancing algorithm to obtain a large number. Also found that, although the lack of convergence iLSTD-greedy

algorithm guarantee, however, when $\lambda \neq 1$, AC-iLSTD-greedy algorithm outperforms AC-iLSTD-random algorithm. For $\lambda = 1$ situation, AC-RLSTD and AC-iLSTD-greedy algorithm performance declined slightly, AC-iLSTD-random algorithm outperforms both above. In the AC-TD any algorithm λ is unstable and, when the worst performance $\lambda = 0.2$ and $\lambda = 0.8$. Therefore, can be drawn from the above analysis, introduce the RLSTD (λ) and iLSTD (λ) algorithm into Critic assessment, AC-

RLSTD and AC-iLSTD algorithm can get better than AC-TD algorithm performance. Although the AC-RLSTD and AC-iLSTD each time step algorithm requires more computation than the AC-TD algorithm, but their data is AC-TD high utilization ratio, this is because the evaluation RLSTD (λ) and the Critic iLSTD (λ) than TD (λ) can make more effective use of empirical data, thus requiring fewer screen or data to learn a better strategy to successfully control the inverted pendulum balance.

a) $\lambda = 0$

b) $\lambda = 0.2$

c) $\lambda = 0.8$

d) $\lambda = 1.0$
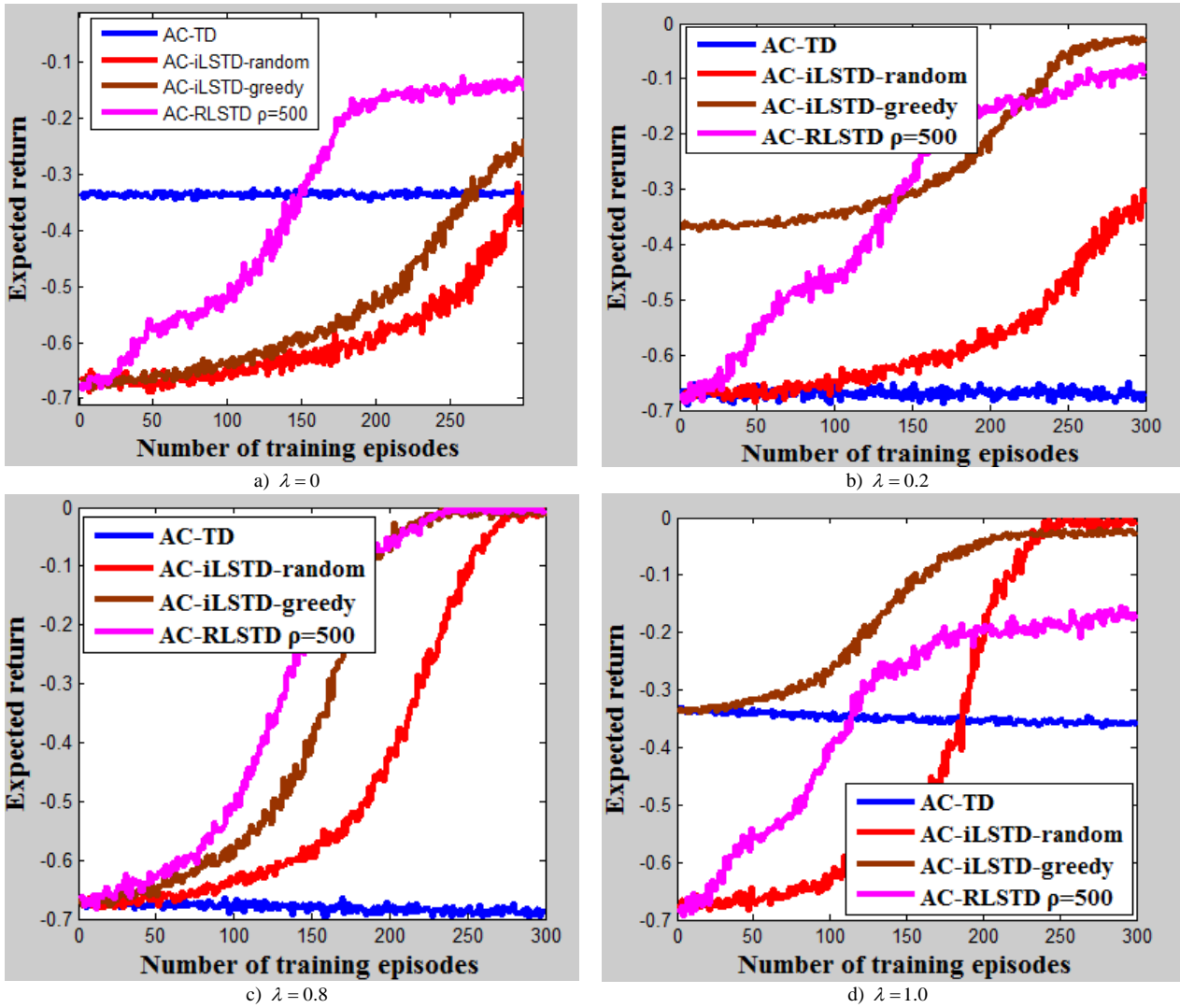
FIGURE 3 Expected returns obtained by different AC algorithms with different λ values

a) $\lambda = 0$

b) $\lambda = 0.2$

c) $\lambda = 0.8$

d) $\lambda = 1.0$
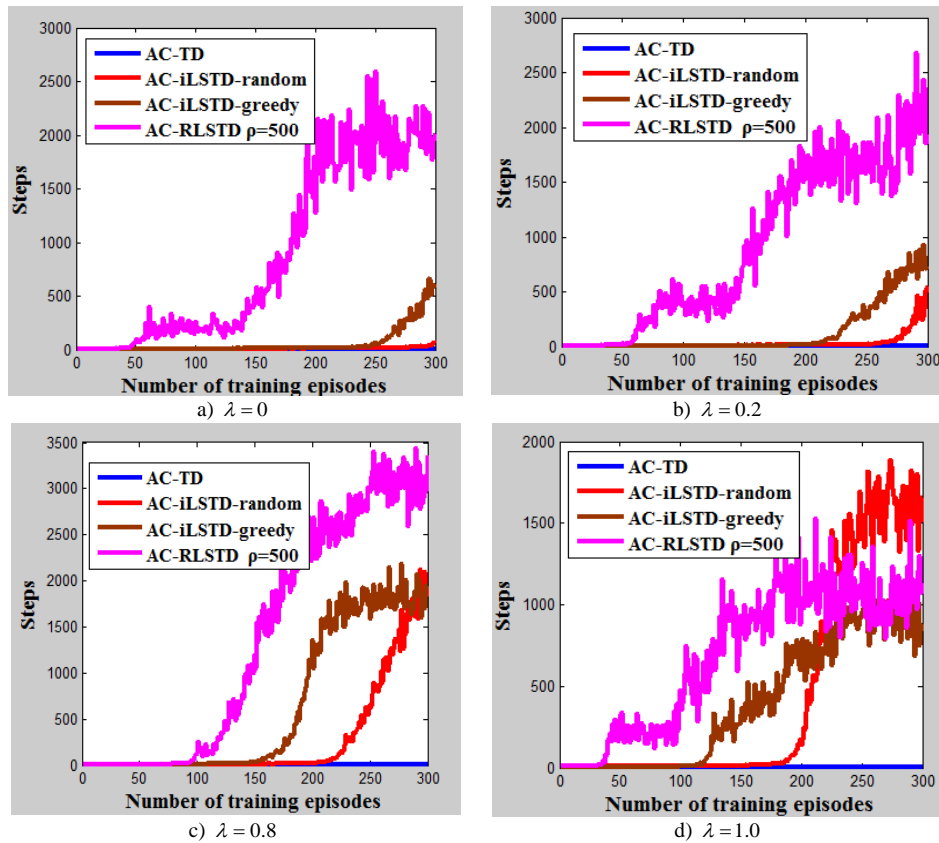
FIGURE 4 Number of balancing steps obtained by different AC algorithms with different λ values

As mentioned in the previous section, the initial value of the initial variance matrix will affect the convergence performance RLSTD (λ) algorithm will RLSTD (λ) algorithm is applied to the AC, AC-RLSTD algorithm above problems also exist. For this reason, the impact of the following test parameters ρ study on the performance of AC-RLSTD. Due to the constant RLSTD larger initial learning performance (λ) algorithm and LSTD (λ) is similar, therefore, has a larger initial constant learning performance AC-RLSTD algorithm perhaps is similar to AC (AC-LSTD) algorithm based LSTD (λ) assessment. To test this possibility, the following also compare the performance AC-RLSTD between the algorithm and the AC-LSTD. Figure 5 shows the performance with different A and B values in the AC-RLSTD algorithm, wherein Figure 5a) and Figure 5b), respectively, by the expected return AC-RLSTD balancing algorithm to obtain the number of steps and the average results of 20 runs. As can be seen, the performance of the algorithm when $\rho = 500$ the AC-RLSTD better than at the time $\rho = 0.1$ and $\rho = 0.5$ .Meanwhile, the simulation results also verified the AC-RLSTD algorithm performance with a larger initial constant A is similar to AC-LSTD. Therefore, when the use of AC-RLSTD algorithm inverted pendulum problem, choose a larger parameter B can obtain a satisfactory performance.

a) Expected returns averaged over 20 runs

b) Number of balancing steps averaged over 20 runs
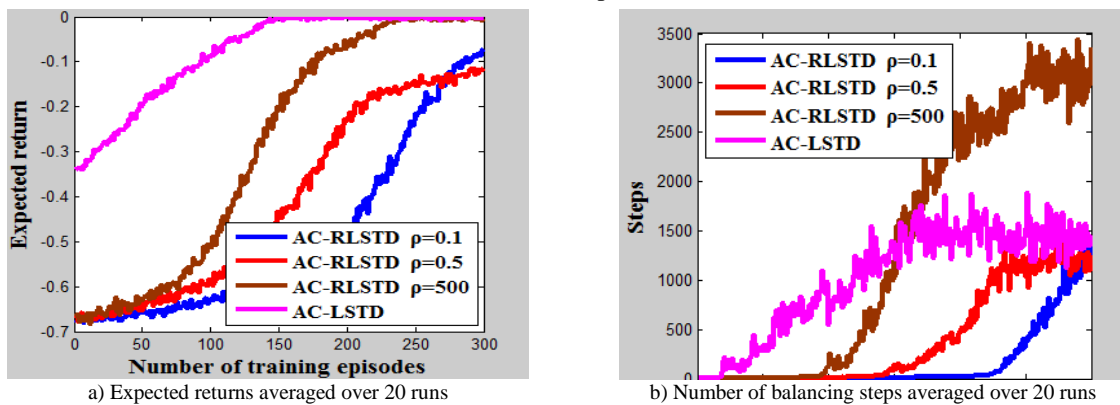
FIGURE 5 Performance comparison of AC-RLSTD algorithm with different initializing constants

6 Conclusions

In summary, compared with the AC-TD method, when $0 < \lambda < 1$, AC-iLSTD and AC-RLSTD algorithm may use the screen or less to obtain a good sample data strategy and their learning better performance. Although iLSTD-greedy algorithm does not guarantee convergence, however, in both learning control problems found AC-iLSTD-greedy algorithm than the AC-iLSTD-random algorithm has better performance. In addition, the simulation also found

that the initial constants ρ has an impact on AC-RLSTD algorithm performance. In the inverted pendulum learning problems, AC-RLSTD use a relatively large value of ρ can get a better performance, while in the car climbing the learning control problems, AC-RLSTD in a relatively small value of ρ can get a better performance.AC-RLSTD algorithm showed two different characteristics according to the recursive least squares method different SNR [13] to explain the situation.

References

[1] Sutton R S, Barto A G 1998 Reinforcement learning: an introduction Cambridge MIT Press
[2] Hirashima Y 2010 A reinforcement learning system for transfer scheduling of freight cars in a train *Proceedings of the International Multi-Conference of Engineering and Computer Scientists* 112-7
[3] Shen Z P, Guo C, Zhang N 2010 A general fuzzified CMAC based reinforcement learning control for ship steering using recursiveleast-squares algorithm *Neurocomputing* **73**(4-6) 700-6
[4] Bradtke SJ, Barto AG 1996 Linear least-squares algorithms for temporal difference learning *Machine Learning* **22**(1-3) 33-57
[5] Boyan J A 2002 Technical update: least-squares temporal difference learning *MachineLearning* **49** 233-46
[6] Geramifard A, Bowling M, Sutton R S 2006 Incremental least-squares temporal difference learning *Proceedings of the Twenty-First National Conference on Artificial Intelligence Boston Massachusetts The AAAI Press* 356-61
[7] Geramifard A, Bowling M, Zinkevich M 2006 iLSTD: eligibility traces and convergence analysis *Proceedings of advances in neural information processing systems Vancouver Canada The MIT Press* 826-33
[8] Lagoudakis MG, Parr R 2003 Least-squares policy iteration *Journal of Machine Learning Research* **4** 1107-49
[9] Bhatnagar S, Sutton R S, Ghavamzadeh M 2009 Natural-gradient actor-critic algorithms *Automatica* **45**(11) 2471-82
[10] Xu X, He H, Hu D 2002 Efficient reinforcementlearning using recursive least-squares methods *Journal of Artificial Intelligence Research* **16** 259-92
[11] Morimura T, Uchibe E, Doya K. 2005 Utilizing natural gradient in

temporal difference reinforcement learning with eligibility traces *Proceedings of the 2nd International Symposium on Information Geometry and its Applications* 256-63
[12] Ljung L, Soderstron T 1983 Theory and practice of recursive identification *Cambridge MIT Press*
[13] Moustakides G V 1997 *IEEE Transactions on Signal Processing* **45**(10) 2468-76
[14] Peters J, Schaal S 2008 Natural actor-critic *Neurocomputing* **71** 1180-90
[15] Bhatnagar S, Sutton R S, Ghavamzadeh M 2007 Incremental natural actor-critic algorithms *Proceedings of Advances in Neural Information Processing Systems Vancouver Canada The MIT Press* 105-12
[16] Sutton R S, McAllester D, Singh S 2000 Policy gradient methods for reinforcement learning with function approximation *Proceedings of Advances in Neural Information Processing Systems* 1057-63
[17] Williams R J 1992 Simple statistical gradient-followingalgorithms for connectionist reinforcement learning *Machine Learning* **8**:229-56
[18] Baxter J, Bartlett P L 2001 Infinite-horizon policy-gradient estimation *Journal of Artificial Intelligence Research* **15**:319-50
[19] Greensmith E, Bartlett PL, Baxter J 2004 Variance reduction techniques forgradient estimation in reinforcement learning *Journal of Machine Learning Reseach* **5** 1471-530
[20] Weaver L, Tao N 2001 The optimal reward baseline forgradient-based reinforcement learning *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence Washington* 538-45

Author

Yang Zhenyu, born on September 1, 1980, Jinan, China

Current position, grades: lecturer.
University studies: computer software.
Scientific interest: machine learning, visualization in scientific computing and artificial intelligence.

Operation Research and Decision Making

Research of deep packet inspection technology in industry control systems based on d-Left counting bloom filter

Kehe Wu, Yi Li*, Long Chen, Fei Chen

Department of Control and Computer Engineering, North China Electric Power University, 102206 Beijing, China

Abstract

Aims at the problem of industry control system logic being tampered by malicious programs, this paper proposed a deep packet inspection model for industry control systems. It adopts d-Left Counting Bloom Filter to implement string matching of characteristic database which has a higher matching efficiency and lower misjudgement rate compared to other algorithms. This model can monitor malicious behaviours in industry control systems effectively and can basically meet the demand of security and reliability in current industry control systems.

Keywords: industry control system, deep packet inspection, d-left counting bloom filter, information security

1 Introduction

With rapid development of computer and network technology, industry control systems have been widely adopted in areas such as electric power, water power, petrochemical industry, medicine, food, automobile and aerospace, and they play a vital role in key infrastructure of a country. Thus how to maintain information security of industry control systems is critical to a country's strategic security [1]. In early days industry control system networks are physically separated to external networks, while in recent years, information technology and the demand of enterprise administration improves rapidly, which make industry control system network and external network integrate smoothly. Especially the adoption of Ethernet and TCP/IP in industry control networks enable these systems communicate with external systems directly, even connect to Internet. While this measure pushes industry production forward extremely, it brings many security problems at the same time like Trojans, viruses and information disclosure and control instruction tampering caused by network attacks [2, 3]. Unfortunately, these problems are not paid enough attention until discovery of Stuxnet in 2010 [4, 5]. After broke into the control system of uranium enrichment factory and nuclear power plant in Iran, Stuxnet tampered the control logic of programmable logic controllers lie in centrifuge control system to make the spin frequency of centrifuge change abnormally, which makes the centrifuge break down ultimately [6-8]. In order to prevent industry control system components from being destroyed by malicious programs, we need to monitor network traffic of industry control system to identify and block malicious behaviours [9, 10].

Aims at the problems discussed above, this paper proposed a deep packet inspection model for industry control systems based on d-Left Counting Bloom Filter. It implements inspection of important field or value in data packets through more detailed rule configuration against specific industry control system protocols. This model can monitor malicious behaviours in industry control systems effectively and improves security and reliability in industry control systems.

2 Deep packet inspection technology

2.1 FUNDAMENTAL OF DEEP PACKET INSPECTION

Deep packet inspection (DPI) is a kind of message inspection technology contrary to traditional packet inspection technology [11]. The so-called "depth" means the level of packets being analysed is deeper than the traditional measures, which only analyse content of IP packets below the Application Layer, including the source address, destination address, source port, destination port and protocol type. However, DPI not only has the function introduced on the front, but also increases the Application Layer analysis which can recognize and content of various applications [12]. Although the industry has not formed a relatively clear definition of DPI technology, it is generally believed that, a device with DPI enabled in addition to the analysis with ordinary message inspection ability (i.e. the message header information, including the source IP address, destination IP address, source port, destination port and protocol type), also can be combined with monitoring between factors such as the payload and the message to implement "deep" recognition [13]. The

* *Corresponding author* e-mail: liyi174748@163.com

fundamental of deep packet inspection is shown in Figure 1.

In this figure, load balancer is responsible for processing data flow through it, and the position it lies is also the position a firewall lies usually. We take data flow as a group of TCP/IP packets, each packet is independent. In order to ensure data packets arrive in correct sequence, the packets should be reordered before data flow

reorganization. Then the load balancer reorganize the data, analyse its protocol in depth, and distribute them to different destination such as Web server, database system, security system, application server etc.. With specific meaning of context, data can be analysed deeply, and can detect attacks from the network layer to the application layer, thus to provide comprehensive protection to computer systems.
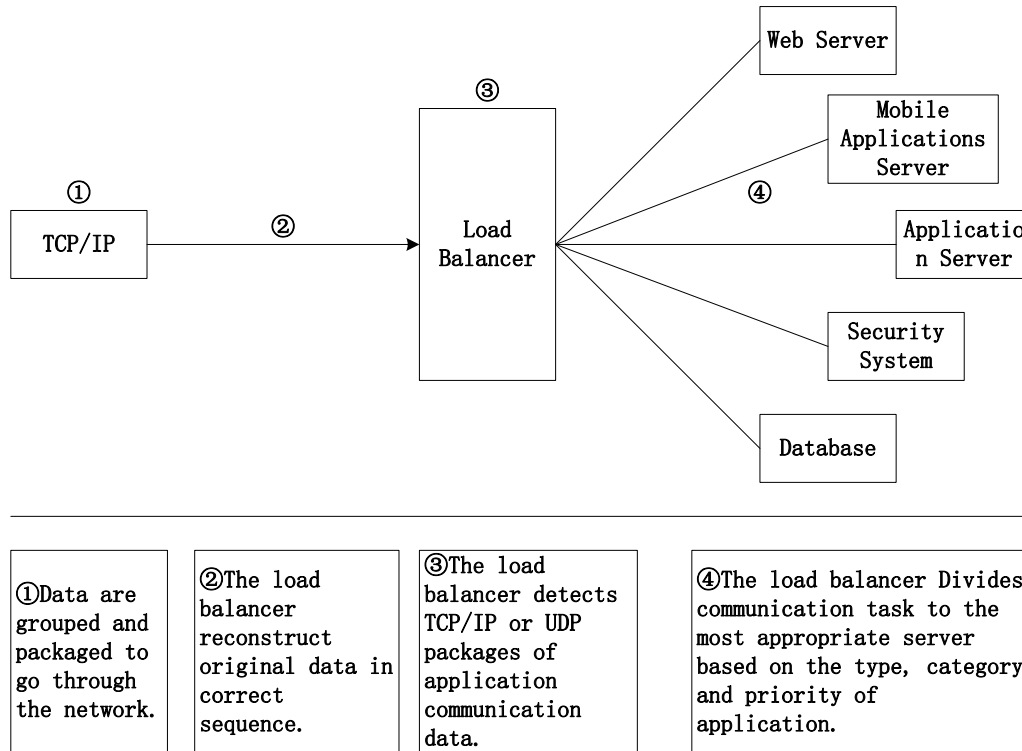
FIGURE 1 Fundamental of deep packet inspection

2.2 ADVANTAGES COMPARED WITH TRADITIONAL PACKET INSPECTION TECHNOLOGY

Although adopts packet filtering to implement packet inspection, but deep packet inspection is different from traditional packet inspection technology, its main advantage is illustrated in the following three aspects.

1) Recognition of protocols in deeper layers.

The traditional packet inspection technology makes the decision of whether forward or block a packet by analyzing its header, for example, it analyses source IP address, source port, destination IP address, destination port, and the protocol used in the IP packet header. If it meets the rules set before, forward it, otherwise it will be dropped. Whereas the deep packet inspection technology is different, it not only can analyse packets on the network layer, but also can analyse packets based on application layer protocols, thus can clearly know what a data bit means in data packets. It is also because of this, a firewall with deep packet inspection can achieve many functions which traditional firewall cannot achieve.

2) Lower Missing Rate

The traditional packet inspection technology analyses and filters separate packet, this approach focus on only part of the overall which makes it easier to miss some attacks, even with state detection. While firewall with deep packet inspection is a real application level firewall, it reorganises packets and can truly realize deep packet content inspection based on state, so it can provide full and complete access control and protection from data link layer to application layer and can reduce missing rate.

3) Stronger Protection Ability

The so-called application layer inspection in some firewalls based on traditional detecting packet inspection just use proxy technology such as URL filtering or other functions in fact, and cannot reconstruct separate packets. For example, if a data flow contains URL bytes is separated into packets, it will not be reverted. That is to say, some malicious programs may bypass this kind of firewalls if skilfully designed. Deep packet inspection technology on the other hand, performs "data reorganization" and "intrusion detection" on the basis of packet filtering, which will be effective to protect computer system from being attacked by application layer attacks [14].

359

3 Analysis of d-Left counting bloom filter

3.1 PROPOSAL OF D-LEFT COUNTING BLOOM FILTER

Bloom Filter is an algorithm Burton H. Bloom proposed in 1970 to solve the problem of whether a certain or some elements are in the set [15]. It uses k independent hash functions to map each element into k bits of a m bits vector. For each element x, the corresponding bit will be set to 1 if the hash function $h_i(x)$ maps into this bit $1 \le i \le k$. If the bit has been set to 1, then the value of this bit won't be changed even it is mapped for the second time. When we query an element y, we use these k hash functions to map it into k bits of the vector. If the value of each bit is 1, we consider this element is in the set, otherwise is not. Bloom Filter is a space saving, efficient data representation and query structure, it uses digit group to represent a collection concisely, and can judge whether an element belongs to the set with a very high probability. However, as hash functions may have collisions, this may cause a fault called "false positive", i.e. it may judge an element in the set which is not in fact. On the other hand, we cannot delete an element from a Bloom Filter, so we cannot perform operations like update and deletion [16, 17].

To solve the problem that elements cannot be deleted in Bloom Filter, Fan L et al. proposed an improved algorithm known as the Counting Bloom Filter (CBF) [18]. It expands the bit vector into a set of counters, and the value of corresponding counter will increase by 1 when it is mapped by a hash function. Conversely the value will decrease by 1 when an element is deleted.

As we know, the counter may overflow and in turn may cause a fault called "false negative", that is to say, it may misjudge an element not in the set. To solve this problem, Bonomi et al. proposed the d-Left Counting Bloom Filter (dLCBF) algorithm [19]. The constructing process of d-Left Counting Bloom Filter is as follows: first use d hash tables, each table have many buckets, each bucket can accommodate a number of (fixed usually) cells, each cell has a fixed number of bits used to hold a fingerprint and a counter to deal with collision. Collision should be dealt with in each table. If there is collision occurs in a hash table, i.e. the same fingerprint is stored in the same location, then just increase the value of the counter by 1. When an element is inserted into the set, we first calculate its hash value and divide it into $d + 1$ segments, in which d of them are taken as store address corresponding to these hash tables, and the other one is the fingerprint to identify the element. Thus, there are d storage locations and a fingerprint. Then we look up these hash tables and choose the leftmost of the lightest load buckets to insert. If the selected position had been stored the same fingerprint, add 1 to the counter of the cell. Conversely in deletion on an element, we get its hash value and find the corresponding counter in the hash table according to the fingerprint and decrease the value of counter by 1. Figure 2 shows the principle of d-Left Counting Bloom Filter algorithm.

Through this algorithm solves the problem of counter overflow may occur in Counting Bloom Filter, it has a flaw may occur in deletion. As we may insert two elements with the same fingerprint and the same bucket index for one of the hash tables, if they are inserted into buckets of different hash tables, when we want to delete one of them, we cannot distinguish which one should be deleted.

To this end, Bonomi et al. introduced a group of random permutations which number is equal to the number of hash tables. After the bucket indexes are divided, we use these permutations to map each bucket index into another value. If we use $H(x) = f(x) = (b, r)$ in which b means bucket index, represent storage location; r means remainder, represent the fingerprint to store. Then the permutations can be represented like this:

$$P_1(f_x) = (b_1, r_1), P_2(f_x) = (b_2, r_2), \cdots, P_d(f_x) = (b_d, r_d)$$

In which $P_1(f_x)$ means storage position and fingerprint of element x for the corresponding hash table. Because permutation means one element will be mapped into another different one, the hash value after the permutation will be different. Thus we can avoid elements from being inserted into the same bucket of a hash table. In process of algorithm realization, linear function can be adopted to realize random permutation. If the range of hash value is 2^q, the random permutation can be written as:

$$P_i(H(x)) = aH(x) \bmod 2^q,$$

in which a is a random odd number in range 2^q.

After that, before insertion, we should look up the selected bucket indexes to find whether they have the same fingerprint, if any increase the counter of corresponding cell by 1. Because of the storage locations of different elements are not coincident except there are collisions appeared. Once collision detection is done before insertion, delete operation will not find two identical fingerprint in different locations. This would solve the flaw when deleting elements.

3.2 COMPARISON BETWEEN DLCBF AND CBF

Now we make a comparison between d-Left Counting Bloom Filter and Counting Bloom Filter. Suppose the set contains m elements, construct a d-Left Counting Bloom Filter like this:

1) the d-Left hash table contains 4 subtables;

2) suppose average load of each bucket is 6 elements ultimately, thus making each subtable contains $m/24$ buckets;

3) each suitable contains 8 cells, that will be able to guarantee the counter not overflow with high probability;

4) each counter of cell contains 2 bits, so it can hold 4 identical fingerprints. The value of counter should be set to 0 at the first time.
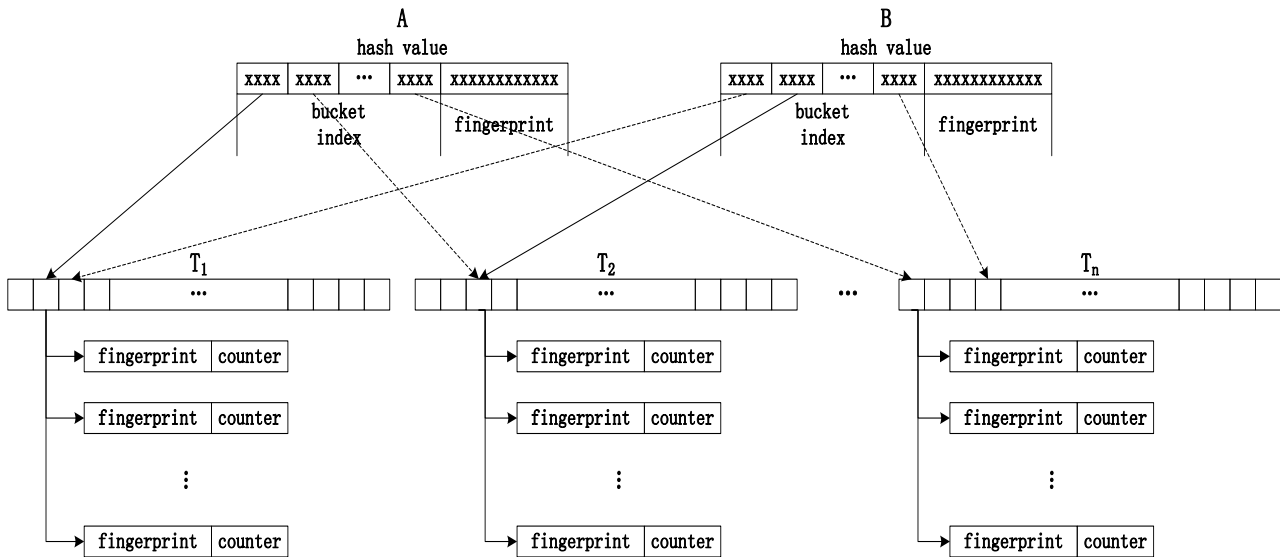
FIGURE 2 Principle of d-Left Counting Bloom Filter

Firstly we compare false positive rate of the two algorithms. Assuming there is a set contains m elements. For Counting Bloom Filter, it is constructed by cm counters, in which c represents ratio parameter between the number of counters n and the number of elements m, each counter occupies 4 bits. When the number of hash functions k in this algorithm adopts the optimal value $c \cdot \ln 2$, its false positive rate is $\left(2^{-\ln 2}\right)^c$, and it occupies $4cm$ bits in total.

For d-Left Counting Bloom Filter, we use r bits to represent the fingerprint. If there is false positive occurs, i.e. there is collision between an element in query and an in the set, the two elements must have the same storage location (the same subtable, same bucket and same cell) and they must have the same fingerprint after hash calculation and permutation. Hence its false positive rate

is $\begin{pmatrix} 4 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 1 \end{pmatrix} \cdot \left(\frac{1}{2}\right)^r = 24 \cdot 2^{-r}$. As each counter have 2 bits, thus each cell occupies $(r+2)$ bits. And because each subtable has $m/24$ buckets, each bucket contains 8 cells, this algorithm needs

$$4 \cdot \frac{m}{24} \cdot 8 \cdot (r+2) = 4m(r+2)/3 \text{ bits in all.}$$

From $4cm = 4m(r+2)/3$ we can acquire when the parameter $c = (r+2)/3$, the bits these two algorithm occupied is equal. Through calculation we can see that when $r \leq 6$, Counting Bloom Filter has a lower false positive rate; whereas when $r \geq 7$, d-Left Counting Bloom Filter is lower and the bits of fingerprint occupied grows, the gap of rate between them is bigger and bigger. Figure 3 illustrates the comparison situation.
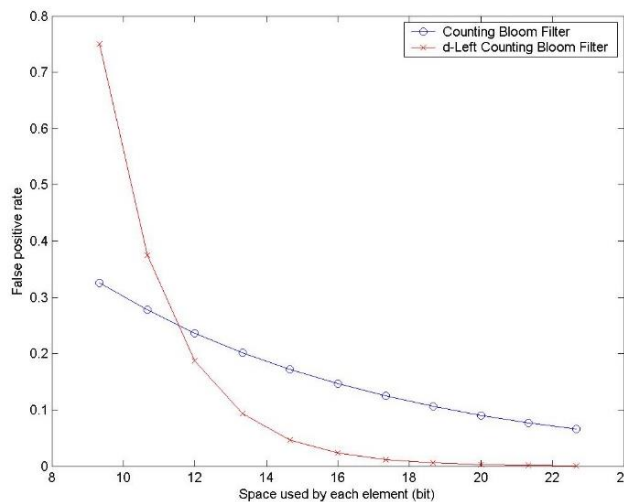
FIGURE 3 Comparison between Counting Bloom Filter and d-Left Counting Bloom Filter

Secondly we compare space occupancy situation of the two algorithms when they have the same false positive rate. Assuming Counting Bloom Filter uses 6 dependent hash

functions, each counter occupies 4 bits. According to Bloom Filter optimal number of hash functions $k = c \cdot \ln 2$, if we want the false positive rate lower than ε, the number

361

of counters n should be at least $k/\ln2 = 8.66$ times more than the number of elements m in set. That is to say, each counter need to adopt 9 counters to guarantee the false positive rate low enough, at this time the Counting Bloom Filter occupies $36m$ bits in total, its false positive rate is 0.01325. For d-Left Counting Bloom Filter, if each fingerprint is represented by 11 bits, it occupies $4m(11+2)/3 = 52m/3$ bits and the false positive rate is 0.01172. The space occupancy ratio of them is $52m/3 \div 36m = 0.48$. In other words, d-Left Counting Bloom Filter achieves a lower false positive rate using just less than half of the space occupied by Counting Bloom Filter.

4 Deep packet inspection model in industry control systems based on d-Left Counting Bloom Filter

This deep packet inspection model is divided into two parts: foreground modules and background modules. The background modules are responsible for recognizing and processing data packets, and recording the condition of handling. Background modules are made up of the characteristic library, traffic recognition module, database module, traffic management module, and log module. The Foreground is responsible for issuing user strategy, and displaying log information [20]. The whole architecture of the model and interactions between each module are shown in Figure 4.
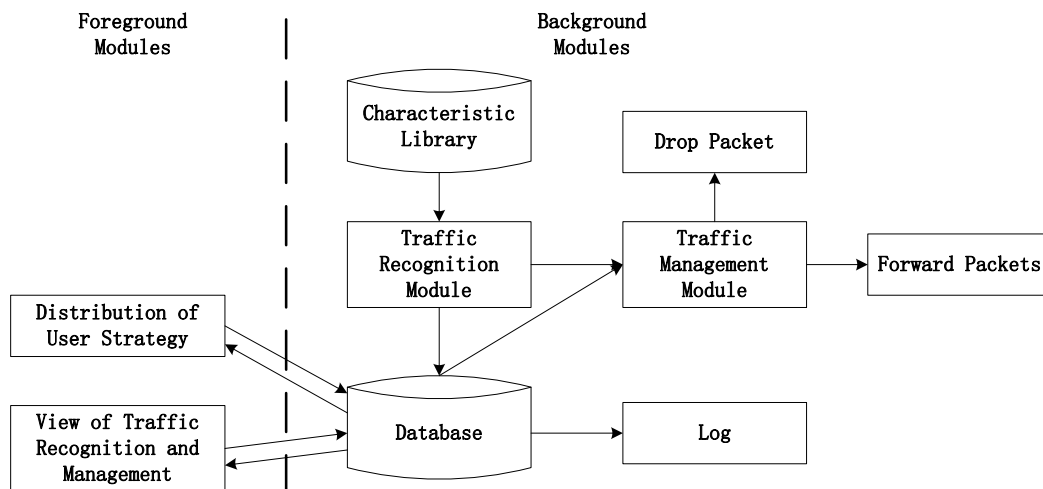
FIGURE 4 Architecture of Deep Packet Inspection Model

The main process of this model is as follows. The packets captured by firewall will be sent to the traffic recognition module then the traffic recognition module loads the access control rules of the characteristics library and perform access control detection to these packets. If they do not meet the requirements of rules the packets will be discarded directly and the recognition results will be stored into database at the same time. Otherwise the packets will be transmitted to the traffic control module, in which recognition strategy is taken out from database module to determine which measure to take on the next step according to the strategy.

4.1 FOREGROUND MODULES

Foreground modules provides display function of the system, mainly including user strategy distribution and viewing traffic recognition and management.

With user strategy distribution module, users can decide IP addresses, protocols, feature strings in packet content and type of control (forward or block). The strategy information will be stored in database after that.

With traffic recognition and management viewing module, users can query traffic recognition results in real time or sometime before. The results also contain traffic amount of corresponding protocol, percentage of total traffic and the type the traffic belongs to etc.

4.2 TRAFFIC RECOGNITION MODULE

In recognition of data traffic, we need to choose different recognition technology according to the characteristic of traffic. To content of some protocols, it can be identified by matching with rules stored in the characteristic library. But to some protocols, we should take other information into consideration like packet size and packet rate, or take load content for deep analysis, hence we should provide specific recognition method for these protocols. According to analysis of industry control system protocols like Modbus and Profibus and some commonly known application protocols, we summarizes the following protocol recognition process, as Figure 5 illustrates [21, 22].

Firstly, we distribute the data packets and check whether they meet the entrance condition, which usually contains common access control parameters like packet size, IP address and port. Then if they meet the entrance condition, we will check whether these packets use known protocols. If they are, the content of packets will be reconstructed and matched with corresponding application protocol rules in characteristic library. If the packets do not

meet entrance condition or they use protocols not known before, the reconstructed content will be matched with specific rules (e.g. feature codes or key words) lie in characteristic library. If the rules are traversed without matching anyone, we consider the recognition results as unknown. Finally, we achieve the recognition results, deposit them into database and transmit them to traffic management module.
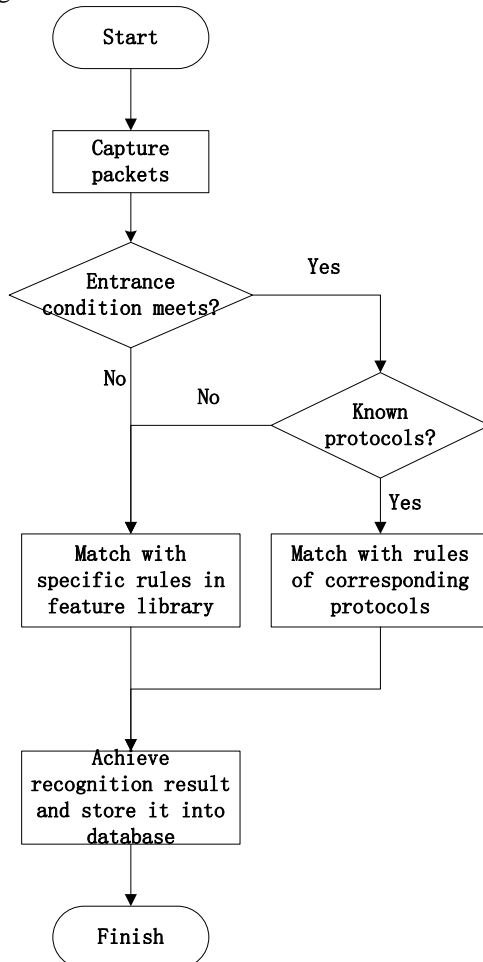
FIGURE 5 Process of Protocol Recognition Module

4.3 TRAFFIC MANAGEMENT MODULE

On system initialization, the traffic recognition module takes user strategies out from characteristic library and put it into a linked list. When there are traffic recognition results achieve, this module determine whether to forward or block the packets according to strategies lies in the linked list. The purpose of doing so is because query in linked list is faster than in the database, but we need to scan the user policy in database regularly, and reload the strategy into the linked list if there are any updates.

4.4 DATABASE

From Figure 4, we can see that the database have interaction with foreground modules, traffic recognition module and traffic management module, in addition it is the data source of the log module. This measure not only avoids much network communication between other modules, but also saves large amount of memory space occupied for data storage.

The database stores various kinds of packet information (e.g. packet size, capture time, header information, load content, recognition result and processing result); the user strategy such as target IP address and protocols. The traffic recognition module deposit recognition results into database for viewing in foreground modules or exporting to log module. Foreground modules put user strategy into database for the use of querying in traffic management module, while foreground modules can view results of recognition and management stored in database, through which avoid direct communication between foreground modules and traffic management module. From this point, the database not only becomes data transmission channel between other modules, but also persistent data.

When there are data packets arrive in traffic management module, if the results of protocol recognition are unknown, it will alert the user to determine whether to forward the packets or block them. Otherwise it will traverse the linked list of strategy, if the strategy corresponding to this protocol order to block this traffic, the packets will be dropped directly; if not, the packets will be forward to user applications. At the same time, the results will be stored to the database; hence can interact between the log module and the foreground module.

4.5 LOG

The function of log module is to record the operation performed by the model and error information, monitor health condition of system, as well as regular export recognition results and condition of blocking from the database for long-term preservation.

5 Conclusions

Focuses on special demand in industry control system application process, through study of deep packet inspection technology and d-Left Counting Bloom Filter in depth, this paper put forward a deep packet inspection model of industry control system based on d-Left Counting Bloom Filter. As d-Left Counting Bloom Filter algorithm has lower space complexity and error rate, it is suitable for implementation in chips; thereby it is convenient for using in high speed integrated circuits integrated into the network equipment, which has a high degree of deployment flexibility with high performance at the same time.

Acknowledgments

References

[1] Berman D, Butts J 2012 Towards characterization of cyber attacks on industrial control systems: Emulating field devices using gumstix technology. In Resilient Control Systems (ISRCS) *5th International Symposium IEEE* 63-8

[2] Shin S 2012 Activities for Control System Security in Japan *In SICE Annual Conference (SICE) IEEE* 667-9

[3] Karnouskos S 2011 Stuxnet worm impact on industrial cyber-physical system security *In IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society IEEE* 4490-4

[4] Cheminod M, Durante L, Valenzano A 2012 System configuration check against security policies in industrial networks. In Industrial Embedded Systems (SIES) *2012 7th IEEE International Symposium IEEE* 247-65

[5] Denning D E 2012 Stuxnet: what has changed? *Future Internet* **4**(3) 672-87

[6] Falliere N, Murchu L O, Chien E 2011 W32 stuxnet dossier *White paper, Symantec Corp Security Response*

[7] Masood R, Um-e-Ghazia U, Anwar Z 2011 SWAM: Stuxnet Worm Analysis in Metasploit *In Frontiers of Information Technology (FIT) IEEE* 142-7

[8] Byres E, Ginter A, Langill J 2011 How Stuxnet spreads–A study of infection paths in best practice systems *White paper*

[9] Faily S, Flechais I 2011 User-centered information security policy development in a post-stuxnet world. In Availability, Reliability and Security (ARES) *2011 Sixth International Conference IEEE* 716-21

[10] Kriaa S, Bouissou M, Piètre-Cambacédès L 2012 Modeling the Stuxnet attack with BDMP: Towards more formal risk assessments. In Risk and Security of Internet and Systems (CRiSIS) *2012 7th International Conference IEEE* 1-8

[11] Riley M C, Scott B 2009 Deep Packet Inspection: The end of the internet as we know it

[12] Building Intelligent Mobile Data Services using Deep Packet Inspection Napatech 2011

[13] Wang J, Wei J, Lin S, Huang W 2013 Research of Deep Packet Inspection System Based on the MapReduce *Journal of Computational Information Systems* **9**(7) 2587-94

[14] Zhou Y 2012 Hardware acceleration for power efficient deep packet inspection *Dublin City University*

[15] Dharmapurikar S, Krishnamurthy P, Sproull T, Lockwood J 2003 Deep packet inspection using parallel bloom filters. In High Performance Interconnects *Proceedings 11th Symposium IEEE* 44-51

[16] Pal S K, Sardana P 2012 Bloom Filters & Their Applications *International Journal of Computer Applications Technology and Research* **1**(1) 25-9

[17] Zink T, Waldvogel M, Scholl M H 2009 Packet forwarding using improved Bloom filters *Master's thesis, University of Konstanz*

[18] Fan L, Cao P, Almeida J, et al 2000 Summary cache: a scalable wide-area web cache sharing protocol *IEEE/ACM Transactions on Networking (TON)* **8**(3) 281-93

[19] Bonomi F, Mitzenmacher M, Panigrahy R, Singh S, Varghese G 2006 An improved construction for counting bloom filters. In Algorithms–ESA *Springer Berlin Heidelberg* 684-95

[20] Byres E, Karsch J, Carter J 2005 NISCC good practice guide on firewall deployment for SCADA and process control networks *National Infrastructure Security Co-Ordination Centre*

[21] Protecting Industrial Control Systems - Recommendations for Europe and Member States ENISA 2011

[22] Samson 1999 PROFIBUS-PA Technical Information, Part 4 *Communication, Frankfurt: Samson AG*

Authors

Kehe Wu, born on August 11, 1962, Jiangsu, China

Current position, grades: professor, Department of Control and Computer Engineering, North China Electric Power University.
University studies: North China Electric Power University.
Scientific interest: intelligent software, cloud computing, information security.
Publications: 80 papers.
Experience: director of the Chinese Association for artificial.

Yi Li, born on October 27, 1988, Shandong, China

Current position, grades: Ph.D. student of North China Electric Power University.
University studies: North China Electric Power University.
Scientific interest: information security, cloud computing.
Publications: 4 papers.

Long Chen, born on March 31, 1988, Shaanxi, China

Current position, grades: Ph.D. student of North China Electric Power University.
University studies: North China Electric Power University.
Scientific interest: cloud computing, distributed storage and networking.
Publications: 4 papers.

Fei Chen, born on April 28, 1986, Jiangsu, China

Current position, grades: Ph.D. student of North China Electric Power University.
University studies: North China Electric Power University.
Scientific interest: electric information security and trusted computing.
Publications: 9 papers.

Operation Research and Decision Making

Study on configuration sequence of indemnificatory community public service facility based on MIV-BP Neural Network

Tianyan Wu*, Jianjun Zhan, Wei Yan

School of Urban Construction and Safety engineering, Shanghai Institute of Technology, Shanghai, China

Abstract

China is seeing large-scale construction of indemnificatory community being. Yet due to lack of dynamic planning and arrangement in advance and little consideration of public service facility configuration sequence, the configuration of public service facility in indemnificatory community is lagging behind and inefficient, failing to attract the residents to move in the community. This paper structures the MIV-BP Neural Network Model, and gives an empirical analysis on the influence sequence of the indemnificatory community public service facilities to the population occupancy rate. The results suggest that the configuration of public service facility in indemnificatory community should be sequentially configured in period and in grade according to the community's specific present situation and developmental conditions as well as the continuous increase of population occupancy rate.

Keywords: indemnificatory community, public service facility, configuration sequence, MIV-BP neural network

1 Introduction

Recent years has seen large-scale construction of indemnificatory communities, such as low-rent housing, affordable housing, and relocation-oriented housing. The traditional indemnificatory community public service facility configuration and planning theory only focuses on the configuration at the end of the planning, without giving consideration to, or before-hand planning and arrangement of, the configuration process of public service facility from the beginning to the end, with little consideration of the sequence of public service facility, blindly follows the plan of public service facility without considering the population change and demands change in indemnificatory community. Consequently, some facilities were constructed prematurely, causing waste due to limited population; some facilities failed to meet the daily fundamental needs of existing residential population and to attract intake population. Therefore, it is high time to study the configuration sequence of indemnificatory community public service facility.

Current researches by domestic and foreign scholars mainly focus on the accessibility and fairness of public service facility configuration. Macintyre S.L [1] holds that the spatial accessibility of public service facility, which is an important indicator of the city residents' life quality, concerns the social fairness and justice of the city public resources allocation. Luo W [2] finds that the accessibility of public service facility is usually positively related to the type of profession and the level of income; individuals with a relatively low social and economic standing usually have a relatively low accessibility to public service facility. Coombes and Jones [3] use questionnaires and field interview to study the accessibility to public service facility by individuals. The accessibility study from the time geography perspective examines the chance of obtaining service based on the behaviour space model and analyses the impact of space distribution of chance on the choice of individual behaviour [4].

A representative scholar of the spatial fairness study of public service facility is Nicholls, who uses the park system of Bryan, Texas, as example and the geography information system technology as platform, analyzes the social and economic attributes of the corresponding regional populations through Mann-Whitney U test, and conducts a comprehensive evaluation of the spatial fairness of the park facilities in Bryan City [5]. Kinman [6] proposes the spatial equality of health care and medical care service facility. Hay [7] believes that urban citizens have equal opportunities of utilizing public service facility. Bach [8] examines the relationship between spatial fairness and benefit of service implementation of public service facility.

Only few scholars have studied the configuration sequence of public service facility. This study uses the MIV-BP Neural Network model to examine the effect of indemnificatory community public service facility on occupancy rate, studies the configuration sequence of indemnificatory community public service facility based on the analysis of the model conclusion and the demand characteristics of public service facility, and is therefore of high practical value.

---

* *Corresponding author* e-mail: wty@sit.edu.cn

2 Demand analysis of indemnificatory community public service facility and construction of MIV-BP neural network model

2.1 DEMAND ANALYSIS OF INDEMNIFICATORY COMMUNITY PUBLIC SERVICE FACILITY

The indemnificatory community public service facility in this study mainly refers to the 11 types of public service facilities defined based on the different functions of the facility according to the Configuration Standards of Public Service Facility in Urban Residential Areas and Urban Residential Communities (DGJ08-55-2006 J100-2006) of Shanghai, including commercial facility, education, culture, fitness, finance, medical care, public administration, municipal public utilities, and greenbelt [9].

Based on the levelled demand theory of Abraham H. Maslow and the questionnaire results, we can divide the demands of indemnificatory community residents for public service facility into several levels, from the low-level fundamental demand to the mid-level comfort demand and the high-level developmental demand. The low-level fundamental demand is the basic subsistence demand for the basic necessities of life, such as food, clothes, housing and transportation, involving the most fundamental demand for daily material and cultural life and the minimum standard of living, such as real property maintenance, cleaning, security, transportation, and water and electricity. The dissatisfaction of these low-level demands, though not affecting the residents' survival, will severely undermine their basic life and therefore require the government and the community to provide corresponding public service facility. The mid-level comfort demand refers to the demand for services for the convenience of residents' life, such as housekeeping, restaurant, shopping, medical care, and children's education. The satisfaction of the mid-level demands improves the convenience level of residents' life and the quality of their material life, which can be realized by the development of public service facility. The high-level developmental demand mainly refers to residents' need for spiritual and cultural life and the demand for community self-governance, including culture, entertainment, and democratic management, which is equal to the high-level demands such as self-respect and self-actualization in Maslow's demand theory. These different levels of demands are summarized in Table 1.

TABLE 1 Demands for indemnificatory community public service facility

Demand Level			Shown As	Corresponding Public Service Facility	
Low-level	Living demands	Fundamental demands	Food and clothes	Business service	Fundamental facility
	Security demands		Utilities	municipal public utilities	
Mid-level	Social intercourse demands	Comfort demands	Healthcare, rest	Medical treatment and public health	Amelioration facility
	Respect demands		Welfare	Community service	
High-level	Self-actualization demands	Developmental demands	Cultural entertainment	Culture and fitness	Advancement facility
			Pursuit for knowledge	Education	

The above table suggests that residents' demands for public service facility evolve gradually from low-level to high-level; when a lower level of demands are satisfied, people would have a next level of demands. In other words, when people's physical demands are met, they will inevitably pursue spiritual demands. The nature of demands gradually evolving has determined that there is a sequence in residents' demands for public service facility.

2.2 CONSTRUCTION AND THEORY OF MIV-BP NEURAL NETWORK MODEL

BP network is a multilayer feed forward network in single-direction dissemination. As shown in Figure 1, it is a neural network of three or more layers, including the input layer, the middle layer (hidden layer), and the output layer, with complete connection between two neighbouring layers and no connection between the neurons of each layer. Luan Qinghua et al. [10] pointed out that the BP neural network features strong mapping ability, flexible self-adapting data processing ability, rapid self-learning ability, and a highly parallel internal link structure. The BP neural network can be regarded as a nonlinear function, in which the network inputs and predictions are the independent variable and dependent variable respectively of the function. When the input node number is N and output node number is M, the BP neural network expresses the functional mapping relationship from N independent variables to M dependent variables. Before it is put into use, the BP neural network must be trained to have associative memory and prediction ability.
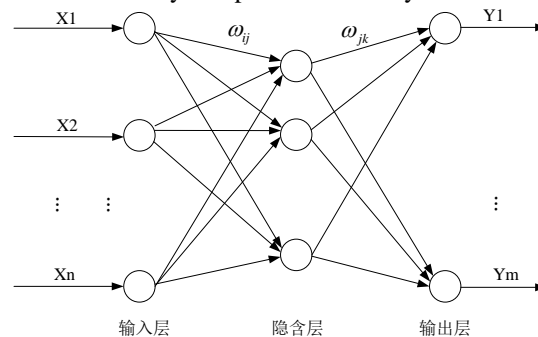
Input layer　　　　Hidden layer　　　　Output layer
FIGURE 1 BP neural network topological structure diagram

In many practical application cases, due to lack of clear theoretical basis, the independent variables, or the network input characteristics, of the neural network can hardly be predetermined. The introduction of some non-important independent variables into the neural network reduces the

precision of the model. Therefore, choosing meaningful independent variables characteristics as the network input data is more often than not a crucial step in analysing prediction problems of using the neural network. To improve the network performance and reduce the errors between prediction output data and expected data, scholars have tried to optimize the network and its input and output data from different perspectives. Yuan Changfeng, Wang Wanlei et al. [11] point out that algorithms such as rough sets algorithm, genetic algorithm, and fuzzy theory can be used to improve the BP neural network training ability. Li Jun et al. [12] introduce the method for accurately predicting film popularity using two indicators by training the neural network through normalization of data in their study of interactive network system. Ding Shifei et al. [13] also hold that the BP neural network based on the genetic algorithm has more powerful local data search capability.

Dombi et al introduce the method of using the Mean Impact Value (MIV) - considered as one of the best indicators for evaluating variable correlation in the neural network - to reflect the variation of weight matrix in the neural network, thereby creating a brand-new thought for solving similar problems [14].

3 Empirical analyses

3.1 DATA COLLECTION

This study uses data from Shanghai Statistical Yearbook, Shanghai Baoshan District Statistical Yearbook, and Questionnaire for an Indemnificatory Community in Baoshan District, Shanghai from 2006 to 2010, as shown in Table 2 and Table 3.

TABLE 2 BP Neural Network Input Data

Year	Public Administration	Fitness	Commerce	Community service	Municipal public utilities	Culture	Education	Medical care	Finance	Greenbelt	Total number of apartments
2006.6	2460	960	1500	60	2300	60	11507	340	804	69839	6563
2006.12	2460	960	1500	80	2680	60	34980	360	804	82804	6563
2007.6	3721	1280	5351	120	2680	200	34980	360	804	88056	12458
2007.12	4432	1280	8927	120	2704	200	34980	4560	804	92414	12458
2008.6	5738	2520	8927	120	3511	1290	34980	4560	804	106601	12458
2008.12	6608	2840	15161	120	3511	2010	34980	4560	804	106601	12458
2009.6	6608	2840	15161	320	3511	2882	34980	4560	804	112850	12458
2009.12	6608	2840	15161	320	3511	3642	34980	4560	804	112850	12458
2010.6	6720	2960	16500	320	3511	3642	34980	4560	804	112850	12458
2010.12	6720	2960	16500	320	3511	3642	34980	4560	804	112850	12458

TABLE 3 BP neural network output data

Year	Number of households
2006.6	236
2006.12	652
2007.6	2018
2007.12	2449
2008.6	3015
2008.12	3314
2009.6	3613
2009.12	3771
2010.6	3841
2010.12	3962

3.2 THE IMPLEMENTATION STEPS OF MIV-BP NEURAL NETWORK

Step 1. Training data selection. From the data shown in Table 2 and Table 3, use the data of the first nine of the ten available time spots as training samples for the neural network and the number of households at the last time spot 2010.12 as the prediction output value of the neural network to test the reliability of the model.

Step 2. BP neural network hidden layer neuron number set. The number setting of neurons in the hidden layer of the BP neural network follows the following principles: First, there need not to be too many hidden layers in the BP neural network; usually one hidden layer suffices to meet the precision requirements. Second, the number of neurons in the hidden layer varies according to different cases; too many neurons are not necessarily good. Based

on tests, at last this study sets one hidden layer, with 10 neurons, in the BP neural network.

Step 3. BP neural network transfer function set. Set the transmission function tansig for the hidden layer neurons, with output as purelin, and use the Levenberg-Marquardt method (trainlm) for the training of the BP neural network.

Step 4. Set the BP neural network training parameters.

To prevent overfitting, set the number of network trainings as 2000 times and choose trainlm training function. Use default values for other parameters.

Step 5. Use the 10th set of data as input and use the trained neural network for prediction to verify the BP neural network effect. Compare the result with the actual value so as to determine the effect of network category prediction.

Step 6. Evaluate using the MIV algorithm.

3.3 THE BP NEURAL NETWORK TRAINING AND EXPLANATION

Neural Network Training, Figure 2a) suggests that the neural network achieves the training effect after five trainings. It is far less than the maximum number we have set and it is the result of normalization processing of selected data that made the training data is more suitable for the neural network training. Figure 2b) and Figure 2c) suggest that the fitting degree R2 = 1 of the neural network training data and the prediction effect of the neural network for the training data is better after training.

a)   b)   c)
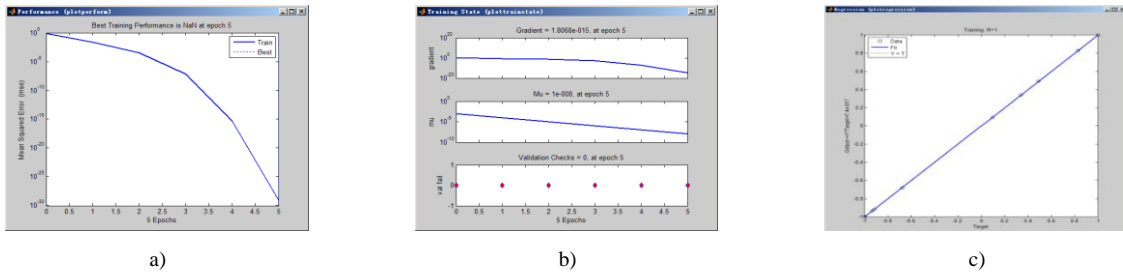
FIGURE 2 a) Neural Network Training Effect, b) Neural Network Training Parameters, c) Neural Network Training Regression.

3.4 MIV RESULT ANALYSIS

Based on the MIV-BP neural network program calculation results, the impact of the public service facility of indemnificatory community on the occupancy rate is shown in Table 4.

TABLE 4 Importance ranking of the impact of public service facilities in an indemnificatory community on occupancy rate.

Variable	MIV	Importance Ranking
Public administration	108.4404	6
fitness	3.093	10
Commerce	373.8412	2
Community service	-35.1358	8
Municipal public utilities	546.009	1
Culture	31.8653	9
Education	85.0184	7
Medical care	-159.2199	4
Finance	-151.5123	5
Greenbelt	273.9467	3

Calculation results suggest that the importance ranking of the impact of indemnificatory community public service facilities on occupancy rate is: municipal public utilities, commercial, and greenbelt facilities of the first group; medical, financial, and public administration facilities of the second group; and education, community service, cultural, and fitness facilities of the third group. In summary:

1) There is a large gap between the cultural and fitness facilities of the community and residents' expectation, lacking cultural exchange centre, indoor fitness gym, and outdoor fitness area; the community service facilities, such as youth activity centre and senior citizen activity centre are also deficient; in terms of education facility, despite the sufficient scale of the fundamental education facility, the teacher resources still lag behind, with a vacuum area in adults education and continuing education. As a result, these facilities become impediments to, instead of the main driving force for, the intake population.

2) The current administrative facilities in the community only meet the ordinary office management requirements of the neighbourhood committee and the property management. Therefore, the financial facility cannot meet the demand and the medical care facility is only limited to the community health centre, unable to meet residents' demand in cases of serious diseases and emergency treatment. Accordingly, these facilities have no significant attraction for the intake population.

3) Municipal public utilities, as fundamental living facility, is the prerequisite for the intake population in the community and ranks the first in the importance ranking, which suggests its crucial effect on the occupancy rate. Although the community business facility includes food market, small supermarket, all types of clothes and beauty salons, food and beverage, and service, due to the limited scale and insufficient number, it has a limited contribution to the occupancy rate of the community despite its high ranking and results in low occupancy rate of communities.

4 Conclusions

The indemnificatory community intake population is mainly mid-income and low-income classes, whose education background, occupational skills, consumption ability, and living standard are relatively low and, compared with the mid-and high-level income classes, are at a relatively low demand level, with the fundamental fund for daily life. In the meanwhile, due to the dynamic and stage nature of the occupancy process, the demand level for public service facility varies at different stages, with different configuration focuses. The dynamic strategy of sequential configuration objective construction for public service facility includes the dynamics of the time dimension and the dynamics of the spatial sequence.

In the time dimension, establish the near-term targets and long-term targets of indemnificatory community public service facility. The near-term targets should have implement stability and operability, giving priority to the configuration of fundamental public service facility to address the existing fundamental living demand for the community residents. The long-term targets should have foresight, giving priority to amelioration and advancement facilities to satisfy residents' increasing comfort and developmental demands. In the meanwhile, during the configuration process, under the influence of the community construction development and residents' demand variation, new facilities should be provided.

In the spatial dimension, establish the target system composed of diversified facilities dominated by public welfare facility. Public welfare facility aims to meet residents' major demand and has strong population aggregation effect, while other types of facilities focus on addressing a specific demand level of the residents.

Therefore, the configuration of indemnificatory community public service facility should adopt by-stage planning, abide by the principles of promoting community

development, accelerating the scale construction of public service facility, and reasonably and sequentially improving public service facility configuration in the middle and long term to determine the public service facility planning objectives, and at the same time improve the public service facility configuration gradually and by stage with full consideration of the operation requirements, economic benefits, and resource sharing. Due to resource limitation, the configuration of indemnificatory community public service facility should first address the problem of "absence", followed by the problem of "deficiency". The public service facility configuration should satisfy residents' different levels of demand in the sequence from the low level to the high level:

1) The near-term planning should focus on fundamental living facility.

The near-term planning objective of indemnificatory community public service facility should focus on the fundamental living service facility for which residents have strong demand for in their daily life and some of the amelioration public service facility for which residents have a pressing demand and control the scale, reserve the land for future demands, leave the room and space for further development, and set the solid schedule. In the meanwhile, it should give consideration to the government-subsidized public welfare facility, such as education and medical care, which have a major impact on population aggregation.

2) The mid-and long-term planning should focus on amelioration and developmental facilities.

The mid-and long-term plan for indemnificatory community public service facility should follow the layout model of balanced configuration and, by evaluating and optimizing the layout models constructed at different stages, obtain the layout model that is more suitable for the indemnificatory community characteristics. In terms of facility configuration, it should make flexible adjustment and supplementation based on the variation of population demands at different stages, so that the balance of layout is reflected both in time and in space, thereby realizing the coordinated development between the public service facility and the intake population scale and demands and satisfy residents' increasing requirements for living standards and self-development. Specially, besides the configuration of fundamental public service facility, a series of amelioration and developmental public service facility should be planned and configured to meet residents' higher level of demand for material and spiritual cultural life. This type of facilities make residents' living conditions more convenient and improve their life value, such as the community service facility of senior citizen recreation service centre, cultural entertainment facility, comprehensive gym, comprehensive stadium, and different types of restaurant and business facilities.

3) A complete public service facility system should be established.

The reasonable planning for the public service facility configuration should be made, identifying the facilities and projects to be configured in the near term, the middle term, and the long term, so as to facilitate the residents that have early moved in, bring into full play the service function and improve the operation efficiency of the facilities. In terms of the spatial configuration, a public service facility network service system should be established with hierarchical functions and spatial layout, more pertinent service, richer contents for each demand level, and greater flexibility.

Acknowledgments

References

[1] Macintyre S L, Macdonald A 2008 *Social Science and Medicine* **67**(6) 900-14

[2] Luo W, Wang F 2003 *Environment and Planning B：Planning and Design* **30** 865-384 (*in Chinese*)

[3] Coombes EG, Jones A P.Hillsdon 2010 *Social Science and Medicine* **70**(6) 816 -22

[4] Neutens T,Versichele M,Schwanen T 2010 *Applied Geography* **30**(4) 561-75

[5] Nicholls S 2001 *Managing Leisure* **6** 201-19

[6] Kim C-W, Lee S-Y, Hong S-C 2005 Equity in utilization of cancer inpatient services by income classes *Health Policy* **72** 187-200

[7] Kinman,E L 1999 Evaluating health services equity at a primary care clinic in Chilimarca Bolivia social *Social science and medicine* **49**(5) 663-78

[8] Hay A M 1995 *Transactions of the Institute of British Geographer* **20** 500-8

[9] Shanghai urban planning administration bureau & Shanghai urban planning & design research institute. Configuration standards of public service facility in urban residential areas and urban residential communities DGJ08-55-2006

[10] Luan Q, Zhu C 2011 *Journal of Software* **6**(12) 2528-34 (*in Chinese*)

[11] Yuan C, Wang W, Lin Y, Chen Y 2011 *Journal of Computers* **6**(4) 776-83 (*in Chinese*)

[12] Li J, Hong S, Xia S, Luo S 2012 *Journal of Network* **7**(12) 2051-6 (*in Chinese*)

[13] Shifei Ding, Xinzheng Xu, Hong Zhu, Jian Wang, Fengxiang Jin 2011 *JWSournal of Computer* **6**(5) 939-46 (*in Chinese*)

[14] Shifeng, Wang X, Yu L, Li Y 2010 Analysis of 30 Cases of MATLAB Neural Network *Beihang University Press Beijing* (*in Chinese*)

Operation Research and Decision Making

Authors

Tian-yan Wu, born in January, 1979, Shanxi Jinzhong, China

Current position, grades: associate professor, PhD.
University studies: Tongji university China.
Scientific interest: the city management, the real estate management, construction project management.
Publications: 20 papers.
Experience: the education experience 8 years, working experience in research – 9 years.

Jian-jun Zhan, born in November, 1989, Jiangxi Fuzhou, China

Current position, grades: graduate student, bachelor's degree.
University studies: Jiangxi university of science and technology, China.
Scientific interest: engineering technology and management.
Publications: 1 paper.

Wei Yan, born in January, 1992, GansuLanzhou, China

Current position, grades: graduate student, bachelor's degree.
University studies: Tianshui normal university, China.
Scientific interest: engineering technology and management.
Publications: 1 paper.

Operation Research and Decision Making

KNN question classification method based on Apriori algorithm

Caixian Chen[1, 2*], Huijian Han[2], Zheng Liu[1, 2]

[1]*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, Shandong, China*

[2]*Shandong Provincial Key Laboratory of Digital Media Technology, Shandong University of Finance and Economics, Jinan 250014, Shandong, China*

Abstract

KNN (K-Nearest Neighbours) algorithm is a classification algorithm that can apply to question classification. However, its time complexity will increase linearly with the increase of training set size, which constrains the actual application effects of this algorithm. In this paper, based on a discussion of disadvantages of traditional KNN methods, an improved KNN algorithm based on Apriori algorithm was proposed. This method extracts the frequent feature set of training samples of different categories and the associated samples. Next, on the basis of correlation analysis of each category of samples, a proper nearest neighbour number k was determined for an unknown category of questions. In the training samples of known categories, k nearest neighbours were selected. And then, according to the category of nearest neighbours, the category of unknown question was identified. Compared with the question classification method of traditional KNN, the improved method could efficiently determine the value of k and decrease time complexity. Our experimental results demonstrated that the improved KNN question classification method improved the efficiency and accuracy of question classification.

Keywords: question classification, KNN, correlation analysis

1 Introduction

In the question and answer system, users can give concise, accurate, and user-friendly answers to questions input by natural language. Answers are generally a length of text. In 1999, a special project of the question and answer system evaluation was introduced in TREC conference. Hence, the question and answer system in the open domain has become a key branch and research focus in the field of natural language processing and information retrieval. Generally, the question and answer system is comprised of three modules, namely, question comprehension, information retrieval and answer extraction. For question comprehension, the recognition of question type is a crucial part, that is, question classification. Question classification is a key factor to locate and test answers and formulate answer extraction strategy.

Currently, the often used question classification algorithms include Native Bayes [4], K-Nearest Neighbours (KNN) [5], SVM (Support Vector Machine) [6] etc. In these algorithms, k-nearest neighbour algorithm is most widely used. However, two issues in KNN algorithm need to be solved: firstly, the way to determine the nearest neighbour number K of samples to be categorized. Secondly, in the classification, the distance between each sample to be classified and all training samples needs to be calculated. Meanwhile, the size of classification samples is often large. To calculate the similarity between thousands of training samples and the

sample to be categorized, the classification performance will be far from satisfactory.

In order to overcome the disadvantages of traditional KNN methods, in this research, an improved KNN algorithm based on Apriori algorithm was proposed. This method extracts the frequent feature set of training samples of different categories and the associated samples. Next, on the basis of correlation analysis of each category of samples, a proper nearest neighbour number k was determined for an unknown category of questions. In the training samples of known categories, k nearest neighbours were selected. And then, according to the category of nearest neighbours, the category of unknown question was identified. Compared with the question classification method of traditional KNN, the improved method could efficiently determine the value of k and decrease time complexity. Our experimental results demonstrated that the improved KNN question classification method improved the efficiency and accuracy of question classification.

2 Related work

KNN is an extension of the nearest neighbour algorithm. Based on the thinking of nearest neighbour, K nearest neighbours of the test samples are selected, and the type of K new nearest samples can be determined. As a non-parameter classification algorithm, KNN has a simple and intuitive principle that is easy to realize. Thus, KNN is widely applied in the field of pattern recognition such as

---

* *Corresponding author* e-mail:chencaixian@163.com

classification and regression [5]. However, two issues in KNN algorithm need to be solved: firstly, the way to determine the nearest neighbour number K of samples to be categorized. According to Bayesian Decision Theory, in order to obtain reliable classification, the larger K is, the better results will be. However, on the other hand, K nearest neighbour samples should be as close to the test samples as possible. Hence, compromises need to be made in reality. The general practice is to determine an initial value first, then keep modifying based on the experimental results and finally reach the optimal value. Many researchers explored into this issue. For example, a comparatively classic reference [7] proposed a K nearest neighbour algorithm that can automatically select the optimal K value. Reference [8] presented a weighted KNN algorithm, which assigns a comparatively large weight to a comparatively nearer neighbour according to the distance from nearest neighbour samples to the test samples. In this way, even K is very large, samples that determine the category of the test sample are nearer samples to it. The weighted method enables the KNN algorithm to be less sensitive to K selection and enhance the robustness of the original algorithm. Secondly, in the classification, the distance between each sample to be classified and all training samples needs to be calculated. Meanwhile, the size of classification samples is often large. To calculate the similarity between thousands of training samples and the sample to be categorized, the classification performance will be far from satisfactory. If some samples in the training set can be reduced before the classification and the final classification accuracy can be ensured, this issue will be solved. Based on this goal, researchers have proposed various approaches to reduce the number of training samples, which can be mainly divided into editing and condensing. Editing methods can remove those samples that may generate classification error or samples surrounded by those samples in other categories, such as references [9,10]. The condensing methods are established based on the following views: samples at the decision boundary are crucial to classification accuracy while samples far from decision boundary impose little impact on the classification. Under the premise of not changing decision boundary, this method removes samples far from the boundary and obtains a comparatively small training set. For these reduction algorithms, commonly used methods include condensed nearest neighbour number rule (CNN) algorithm proposed by reference [11] in 1968. This algorithm can effectively reduce the size of training set, but often retain some samples far from classification boundary. In reference [12], a condensing method based on Voronoi diagram was proposed. The condensing set obtained from this algorithm not only accurately classifies training samples but also generates a classification boundary for all training samples. However, since the Voronoi diagram is introduced, the complexity of this algorithm is considerable. Reference [13] proposed a Decremental Reduction Optimization Procedure 1

(DROP1), and a series of improved algorithms on this basis, including DROP2 and DROP5. Subsequently, reference [14] presented the Improved KNN (IKNN). Through repeated iteration, this algorithm reduces most samples in the training set that cannot match the sample to be tested. This algorithm especially applies to the circumstance of high sample feature dimension. In addition, reference [15] proposed a Template Reduction for KNN (TRKNN), which defines a nearest neighbour chain table. Based on the table, the training set can be divided into the condensing set (generally comprised of samples near classification boundary, i.e. the new training set) and reduced set (generally the internal samples). In addition to the above algorithms, references [16, 17] presented a condensing algorithm based on other principles. Reference [18] presented a mixed model algorithm that combines editing and condensing. Based on KNN, our research proposed a K nearest neighbour algorithm based on Apriori algorithm.

3 Question classification background

3.1 QUESTION CLASSIFICATION AND PROCESS

Question classification is an instructional learning process. It identifies the relation model between question features and question category based on a classified training question set. Next, the relation model from the instructional process can determine the category of new question. Let us set a group of conceptual question C and a group of training question Q. Conceptual questions and questions in the question base may satisfy the hierarchical relation h of a concept. There is also a target concept T:

$$T : Q \rightarrow C . \tag{1}$$

T maps a question case to a category. For question q in Q, $T(q)$ is known. By instructing the study of the training question set, we can find a model H similar to T:

$$g_j(x) = \arg_i \max(k_i) . \tag{2}$$

For a new question qn, $H(qn)$ indicates the classification results of qn. The establishment of a classification system or the objective of classification study is to identify a H most similar to T. In other words, when an evaluation function f is given, the goal of study should enable T and H to meet:

$$\min \left(\sum_{i-1}^{|D|} f \left(T(d_i) - H(d_i) \right) \right) . \tag{3}$$

Generally, question classification needs to settle 5 problems:

1) To acquire the training question set: whether the training question set is properly selected imposes remarkable impact on question classifier. The training question set should be able to represent question in each category to be processed by the classification system. In

general, the training question set should be widely recognized corpus through manual sorting.

2) To establish question representation model: to select language elements (or question features) and mathematical forms to organize these language elements to represent questions. This is a key technical issue in question classification. At present, most question classification methods and systems adopted characteristics or phrases to represent language elements of question semantics. Representation models mainly include Boolean model and vector space model.

3) To choose question features: language is an open system so that the digitalized question as language should be open. Its size, structure, language elements and the information contained in the question are open, and characteristics of the question are not limited. The question classification system should select as few question features as possible accurately and closely related to the question theme for question classification.

4) To select the classification method: the selection of methods to establish the mapping relation from question features to question category is a core issue of question classification. Commonly used methods include Native Bayes, KNN, class-centre vector, regression model, and Support Vector Machine and so on. In fact, KNN method and Support Vector Machine method are often used, which present efficient classification effects and remarkable stability.

5) Performance evaluation model: ways to evaluate classification methods and system performance. The performance evaluation model that truly reflects the internal characteristics of question classification can be used as the target function to improve the target function of classification system. In the question classification, the selection of valuation parameters is subject to the specific classification question. Single-label classification question (one test question only belongs to one category) and multi-label classification question (one test question can belong to several categories) adopt different evaluation parameters. Currently, commonly used classification performance evaluation indicators include recall ratio and precision ratio, which originate from two terms in information retrieval.

FIGURE 1 Question classification process

As shown in the Figure 1, feature selection, classification training and testing constitute a loop. According to the test results, parameters of feature

selection and classification training are adjusted so that the classifier can reach the optimal classification effects.

3.2 INTRODUCTION OF TRADITIONAL OF KNN

KNN is a question classification method based on vector space model. Assume x is a question, and its vector model is $x = (x_1, x_2, ..., x_n)$. Each dimension of question vector x corresponds to each word of question representation, also known as attribute. $C_i = (x_1^i, x_2^i, ..., x_{ni}^i)$ is a question category with class identifier containing question $x_1^i, x_2^i, ..., x_{ni}^i$. The number of questions is ni. Let there be m question training classification questions $C_1, C_2, ..., C_m$, the classification process of KNN is as follows: for a given test question x, in the training question set of all categories $C_1, C_2, ..., C_m$, the similarity between two questions can be used to find $k (k \geq 1)$ nearest training questions. The number of questions of Category C_i was $k_i (i = 1, 2, ..., m)$ and:

$$\sum_{i=1}^{m} k_i = k . \tag{4}$$

The discrimination function of two commonly used question classification is as follows:

Discrimination function 1: (Let $g_i(x) = k_i$, $i = 1, 2, ..., m$ and $X \in C_j$) is decided according to Equation (5):

$$g_j(x) = \arg_i \max(k_i) . \tag{5}$$

In other words, the determination of category of test question adopted the relative majority vote method. In other words, among the category of K nearest neighbour question, the category containing the most questions is taken as the category of ultimate test questions.

Reference [19] presented the weighted KNN decision rule as:

$$score(d, C_i) = \sum_{d_j \in KNN(d)} sim(d, d_j) \delta(d_j, C_i), \tag{6}$$

where $sim(d, d_j)$ is the similarity between d and d_j (the samples to be tested), KNN(d) is the k nearest neighbour set of question d and $\delta(d_j, C_i)$ is used to indicate whether question d_j belongs to C_i:

$$\delta(d_j, C_i) = \begin{cases} 1, & d_j \in C_i \\ 0, & d_j \notin C_i \end{cases} . \tag{7}$$

Methods to calculate the question similarity include Euclidean distance, vector inner product and included angle cosine. The included angle cosine is often used to calculate the similarity between questions. The equation is as follows:
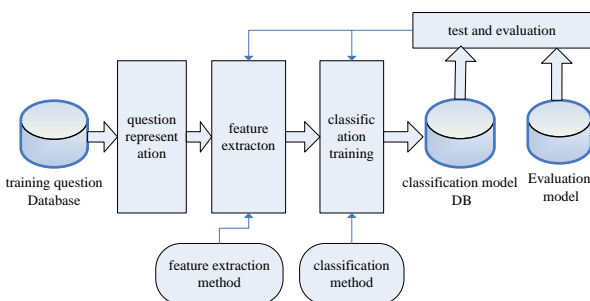
$$sim\left(d_1,d_2\right)=\frac{\sum_{i=1}^{n}W_{1i}W_{2i}}{\sum_{i=1}^{n}W_{1i}W_{2i}} \qquad (8)$$

where W_{1i} and W_{2i} represented the weight of the i-th feature word in question vector of question d_1 and d_2. The greater the cosine is, the more similar the two questions become, and the more likely the question represented by two vectors may belong to the same category, and vice versa.

The question d to be tested belongs to the category with the highest $score\left(d,C_i\right)$. X is used to replace d and X_1^i is used to replace the first question of C_i category in k nearest neighbours. The actual value of $\delta(d_j,C_i)$ was substituted into it. And the following decision function was obtained:

Discrimination function 2: let

$$f_i\left(X\right)=\sum_{l=1}^{k_i}sim\left(X,X_l^i\right),i=1,2,...,m\,, \qquad (9)$$

$X \in C_j$ was determined by Equation (10):

$$g_j\left(x\right)=\arg_i\max\left(g_i\left(X\right)\right). \qquad (10)$$

4 Improved KNN question classification algorithm

In the implementation process of traditional KNN algorithm, the distance between tested question and each training question must be calculated. Besides, the nearest neighbour number k cannot be determined. This influences the promotion of KNN algorithm classification. This paper proposed an improved KNN-based question classification method using the correlation analysis.

4.1 CONCEPTUAL FRAMEWORK OF
ASSOCIATION RULES

Let item set $I=\left\{i_1,i_2,...,i_m\right\}$ be the set of m different symbols. Each symbol was called an item. D is a set comprised by several transactions T. T is the subset of transaction set I. Each item has the unique identifier TID. If X is a sub-set of T, T contains item set X.

Definition 1, item set: set of items; the set including k items is known as k-item set.

Definition 2, supportive number of item set: the number of items in D is regarded as the supportive number of item set X, which can be expressed as:

$$support\_num(X)=\left|\left\{T\mid T\in D,X\subseteq T\right\}\right|.$$

Definition 3, item set support degree indicates the probability of item set occurred in D.

Definition 4, frequent item set: the item set with the minimum support threshold larger or equivalent to that designated by users.

4.2 BASIC IDEA OF THE ALGORITHE

Let there be two categories, P and Q, and 4 feature words, p1, p2, q1 and q2. 4 feature words can produce 15 kinds of non-empty sets {p1}, {p2}, {q1}, {q2}, {p1, p2}… {p1, p2, q1, q2}. Apriori algorithm was used to record question objective that contains the characteristic set under different circumstances. For example, there are 6 objectives that contain the feature word p1, and 2 objectives that contain the feature word p1, p2 and q1. In this way, the frequent item set of feature word was established. When classifying the classification question (p1, p2, q1), we can directly find the question corresponding to the frequent item set of feature word {p1, p2, q1}. As an initial nearest neighbour of classification question, the nearest neighbour number K of the questions to be categorized could be determined ultimately according to the number of initial nearest neighbours. Next, the similarity of questions can be calculated. According to the categories of the first K nearest neighbours, the category of questions to be classified can be determined. For example, the square category in the Figure 2 made for 2/3. Thus, circular objectives were determined as the square category.
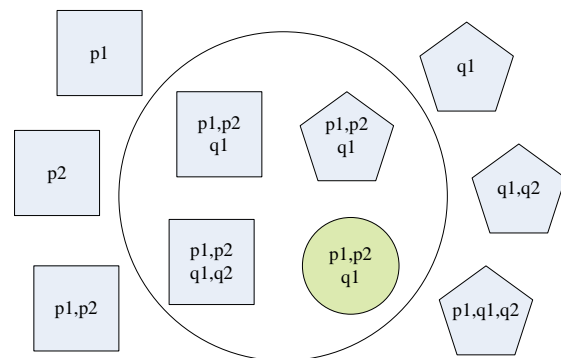
FIGURE 2 Diagram of improved KNN question classification algorithm

It can be seen that the improved classification algorithm could efficient reduce the time complexity of algorithm. However, at the earlier stage, a mapping relation between feature word frequent item set and associated training question objective set can be established, which requires certain time-space consumption.

4.3 QUESTION SIMILARITY CALCULATION
METHOD

Methods to calculate the question similarity generally include the Euclidean distance, vector inner product and included angle cosine. In order to enhance the accuracy of question similarity calculation, this research applied HowNet to adopt semantics-based question similarity calculation method.

HowNet is a repository with the description objective of concepts represented by Chinese and English terms, which can reveal the relationship between concepts as well as between the basic attributes of concepts. HowNet

describes the Chinese terms based on the conceptual framework of "sememe". Sememe can be regarded as the smallest and fundamental Chinese semantic unit that is not easy to divide. Since Chinese words express different meanings in different contexts, Hownet comprehends Chinese words as a set of several sememes. Each entry of the semantic dictionary of HowNet is comprised of a sememe of a word and its description. In other words, one entry corresponds to one sememe of a word. Meanwhile, each sememe is described by several sememe. HowNet provides a classification tree of sememe, and there is a hyponymic semantic relation between parent-node and child-node. Thus, the classification tree of sememe can be used to calculate the semantic similarity between two words.

4.3.1 Calculation of sememe similarity

$$sim(p_1, p_2) = \frac{2 \times Spd(p_1, p_2)}{Depth(p_1) + Depth(p_2)}, \qquad (11)$$

where p_1 and p_2 represent two sememe sources, and $Spd(p_1, p_2)$ shows the coincidence degree of p_1 and p_2. $Depth(p)$ is the depth of a sememe in the sememe tree.

4.3.2 Similarity calculation of notional words

In HowNet, the concept of content words (sememes) can be divided into 4 parts:
1) primary fundamental sememe description: the first sememe in DEF item;
2) description of other fundamental sememe: all the other independent sememe or specific words in DEF item;
3) description of relational sememe: in DEF item, "relational sememe =fundamental sememe " or "relational sememe = (specific words) " or " (relational sememe =specific words)" to describe the notional part;
4) description of symbol sememe: in DEF item, "relational symbol fundamental sememe " or "relational symbol (specific words)" to describe the notional part.

Thus, the similarity corresponding to 4 parts of the 2 notions is respectively marked as $sim_1(C_1, C_2)$, $sim_2(C_1, C_2)$, $sim_3(C_1, C_2)$ and $sim_4(C_1, C_2)$. In this way, the entire similarity of notional words is:

$$sim(C_1, C_2) = \beta_1 sim_1(C_1, C_2) + \sum_{i=2}^{4} \beta_1 \beta_i sim_i(C_1, C_2), \quad (12)$$

where β_i satisfies:

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1, \beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4 \geq 0.$$

4.3.3 Similarity calculation of Chinese words

For two Chinese words W_1 and W_2, if W_1 has n sememe items: $c_{11}, c_{12}, ..., c_{1n}$ and W_2 has m sememe items $c_{21}, c_{22}, ..., c_{2m}$. Let the similarity between W_1 and W_2 be the maximum similarity of each sememe items. Then we have:

$$sim(W_1, W_2) = \max_{1 \leq i \leq n, 1 \leq j \leq m} sim(C_{1i}, C_{2j}). \qquad (13)$$

4.3.4 Similarity calculation of Chinese sentences

For two sentences S_1 and S_2, S_1 has n words: $w_{11}, w_{12}, ..., w_{1n}$. S_2 has m words: $w_{21}, w_{22}, ..., w_{2m}$. The calculation method of sentence similarity is: based on the word set of two sentences, one word is selected from one set to calculate the similarity with each word in another set. The word pairs of the maximum similarity are selected. The loop is not stopped until the first set word is empty. Next, the similarity of selected word pairs is added and then divides the word number contained in the first set. Finally, the calculation results based on two sets are averaged to obtain the similarity of two sentences. The calculation equation is as follows:

$$sim(S_1, S_2) = \left. \left(\frac{\sum_{u=1}^{n} \max_{1 \leq v \leq m} sim(w_{1u}, w_{1u})}{n} + \frac{\sum_{v=1}^{m} \max_{1 \leq u \leq n} sim(w_{1u}, w_{1u})}{m} \right) \middle/ 2 \right. . (13)$$

4.4 REALIZATION OF ALGORITHM

Our research proposed a KNN question classification method based on Apriori algorithm. With correlation analysis, the nearest neighbour was selected, which avoided defects of traditional KNN. Compared with traditional methods, our method effectively improved time complexity and *K* selection. The improvement methods are mainly divided into two stages:
1) based on correlation analysis, the frequent feature word set and the associated training question were extracted;
2) based on findings of correlation analysis, the initial nearest neighbour of questions to be classified was determined, as well as the final nearest neighbour number K. Next, KNN was used for question classification.

1) Extraction of frequent feature word set and associated training question based on correlation analysis

a) Let the total number of question category be *m*, and the category is $C_1, C_2, ..., C_m$. The number of training samples of each category is noted as $N_1, N_2, ..., N_m$; questions in the training set are processed in advance. With $\chi2$ statistical approach, a certain amount of questions of different categories in the training set are selected, and noted as the feature word of N_f (for example, 10 characteristics are selected from each category);

b) Scan all training questions and express each question as $m \times N_f$ -dimensional question vector comprised by feature word of all categories. TF-IDF is used to calculate the characteristic weight;

c) Extract the frequent feature set and the associated questions from each category; this step only considers the characteristics of the category of each training question, and the rest will be omitted for now; each category is processed respectively, including the following procedures:

i) each question of this category is seen as a single transaction, and the included characteristics of this category are seen as the data item of transaction. Item set is also a feature word set of this category. The minimum supportive degree is set and Apriori algorithm is used to enable the question category to meet all item sets of minimum supportive degree threshold, that is, the frequent item set of this question category is produced;

ii) for each frequent item set, the associated training question is preserved, and the training question that contains all features of a frequent item set is the associated training question of this frequent item set;

2) With the findings of correlation analysis, the initial nearest neighbour of the question to be classified is determined as well as the ultimate nearest neighbour number K. The question classification is conducted based on the category of nearest neighbour.

a) For a question to be classified, the pre-treatment is conducted. Next, the extracted feature word of each category can be used to represent this question and to obtain $m \times N_f$ -dimensional question vector. TF-IDF is also applied to calculate the characteristic weight;

b) The feature word weight of each category in question vector of a question to be classified is respectively summed up and arranged in a declining order. Categories in the top 3 are selected and noted as C_x, C_y, C_z and the characteristics;

c) Feature words belonging to the top 3 categories as obtained from *ii* are selected. The maximum frequent item set is found in the corresponding category to acquire the associated training question. These training questions are used as the initial nearest neighbours for a question to be classified. Let the associated training question set be S_x, S_y and S_z respectively. The number of questions is N_x, N_y and N_z. Let $k = \min\left(2.5 \times N_x, N_x + N_y + N_z\right)$;

d) The cosine similarity between the question to be classified and each initial nearest neighbour question is calculated;

e) The similarity is arranged in a declining order, and the first *k* training questions are selected. The file number of 3 categories is noted and the similarity is accumulated according to different categories. Thus, the average similarity between the question to be classified and the nearest neighbour question of each category. The category with the maximum mean is determined as the category of the question.

5 Experiment

5.1 DATE SET

For Chinese question classification, there is no uniform standard question test set and training set so far. In our research, the proposed question set is a set of sentences of demarcated question type by a certain question classification system. Since the addition of QA test tasks in TREC-8, TREC conference provides masses of English question sets for QA evaluation in an annual fashion. Thus, the free question set of TREC2013 was used in our experiment, which was adopted into a part of the Chinese question set by translating some questions and transforming some questions. In addition, some questions were extracted from the previously developed recruitment question and answer system of Shandong Economic University. Besides, some questions were collected from the Internet. Both formed the Chinese question set in the context, and a total of 1500 questions were included.

5.2 RESULT ANALYSIS

Evaluation of a classifier is a key research topic in question classification. For different goals, researchers have proposed many evaluation approaches for question classification such as: recall ratio, precision ratio, F1 test value, and macro-averaging, micro-averaging and so on. In our research, the classic information retrieval evaluation criteria of recall ratio, precision ratio and F test value are used for our evaluation: recall ratio indicates the ratio between accurately identified sample size by the classifier and the sample size belonging to this category. The accuracy rate indicates the proportion of samples really belonging to this category among samples classified in this category by the classifier. The mathematical equation is:

$$R = \frac{A}{A+C}, \tag{14}$$

$$P = \frac{A}{A+B}. \tag{15}$$

A represents the number of questions that belong to the category in the manual sorting criteria and are actually classified by the classifier in this category. *B* represents the number of questions that do not belong to the category in the manual sorting criteria but are actually classified by the classifier in this category. *C* represents the number of questions that belong to the category but are distributed in other categories; *D* represents the number of questions that do not belong to the category and are not classified in the category, as shown in Table 1:

TABLE 1 Meaning of ABCD

	Number of questions truly belong to this category	Number of questions truly not belong to this category
Number of questions belong to this category	A	B
Number of questions not belong to this category	C	D

For a certain category, recall ratio and precision ratio reflect two aspects of the classification quality. Recall ratio and precision ratio are mutually influenced. Under normal circumstances, precision ratio will decrease with the rise of recall ratio, and it is hard to ensure that both are high. Thus, in order to comprehensively reflect the performance

of classification system, recall ratio and precision ratio should be comprehensively considered. The integration of the two indicators will produce a new indicator of evaluation - F1 test value. The mathematical equation is as follow:

$$F1(P,R) = \frac{2 \times P \times R}{P + R} \ . \tag{16}$$

In order to avoid the influence of a small size of sample on the test results, the cities, dates, specific Figures 3-5, the total number of money, the definitions are selected because they contain 381 questions for the test. The results are shown in Table 2.

TABLE 2 Comparison of experimental results

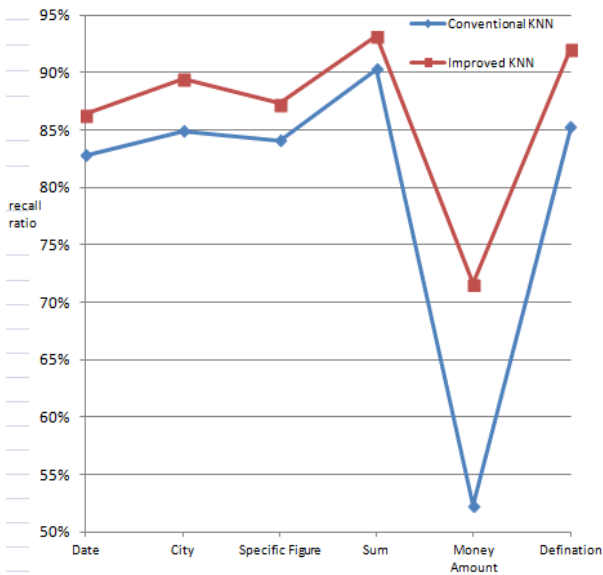| Classification Method	Conventional KNN			Improved KNN		
Question type	precision ratio	recall ratio	F1 test value	precision ratio	recall ratio	F1 test value
Date	87.4	82.8	85.0	93.8	86.4	89.9
City	84.8	84.9	84.8	83.6	89.5	86.4
Specific Figure	87.1	84.1	85.6	90.9	87.3	89.1
Sum	55.6	90.3	69.3	66.2	93.2	77.4
Money Amount	78.2	52.3	62.7	89.4	71.6	79.5
Definition	83.7	85.3	84.5	89.5	92.1	90.8

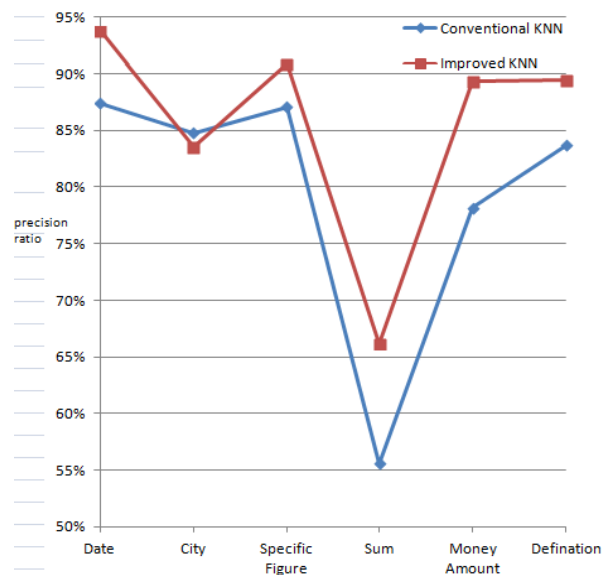FIGURE 3 Recall ratio contrast

FIGURE 4 Precision ratio contrast
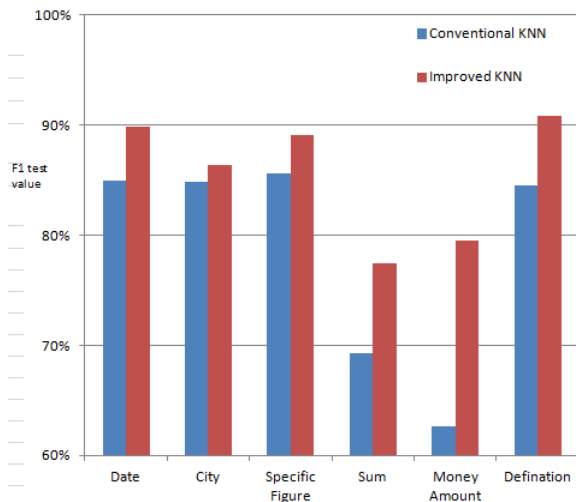
FIGURE 5 F1 test value contrast

It can be observed from the experimental results of traditional KNN algorithm that the recall ratio of money amount is comparatively low, and the precision ratio of the total number is low. It should be noted that the training question number of the category of total number is larger than other categories, and the money amount features share similarity with that of the total number. These led to the mis-categorization of money amount questions in the category of the total number. Meanwhile, the improved KNN algorithm conducted correlation analysis for question vectors in each category and applied the average similarity between the question to be classified and nearest neighbour samples of each category in category determination. This improved algorithm effectively reduced the mis-categorization of text category.

The above experiments presented that feature word from correlation analysis and related information between training questions can be feasibly used to select the nearest neighbour of nearest question. Compared with the traditional KNN method, the calculation load of selecting nearest neighbour can be substantially reduced to about 1/3 of time complexity of traditional methods; meanwhile, nearest neighbours in training samples of different categories can be found via the frequent feature word set. The nearest neighbour number in different categories is of great reference values for determining the nearest neighbour number k of test question. The experimental results presented that the proposed nearest neighbour

selection approach effectively reduced the involvement of training questions with limited similarity in the category judgment of test question. Meanwhile, as frequent item set was produced by Apriori algorithm and the minimum supportive degree was set, the proposed method is not very sensitive to the local features of samples so that it can improve the accuracy of the classification to some degree.

6 Conclusions

Our research proposed an improved KNN question classification method based on Apriori algorithm, which modified the determination method of nearest neighbour number k and reduced the time complexity of classification. The experiment demonstrated that compared with traditional methods, time complexity of the improved KNN question classification method was relatively smaller and the classification accuracy was high. Of course, this method presents certain disadvantages: firstly, correlation analysis using Apriori algorithm of various categories of questions causes remarkable consumption temporally and spatially. Secondly,

For a test question, the retrieval of frequent item set that meets the minimum supportive degree cannot accurately find all the neighbour question vectors. In the experiment, about 80% nearest neighbours of test question can be found. In addition, if the characteristic of a category is not remarkably obvious, this will result in the failure of or greatly limited extraction of associated information that abides by the minimum supportive degree. The solution of the question needs further improvement of Apriori algorithm to extract the associated information, thus improving the recall ratio of test question nearest neighbour.

References

[1] Cai L, Zhou G, Liu K, Zhao J 2011 Learning the Latent Topics for Question Retrieval in Community QA *Proceedings of 5th international joint conference on Natural Language Processing* Chiang Mai Thailand 273-281
[2] Zhou T C, Lin C Y, King I, Lyu M R 2011 Learning to Suggest Questions in Online I:orums *Proceedings of the 25th MI Conference on Artificial Intelligence* San Francisco CA USA 1298-303
[3] Ji Z, Xu F, Wang B 2012 Question-answer Topic Model for Question Retrieval in Community Question Answering. *Proceedings of the 21st ACM international conference on Information and knowledge management* Maui HI USA 2471-4

[4] Gong X, Sun J, Shi Z 2002 The Classifier of Initiative Bayesian Network *Journal of Computer Research and Development* **39**(5) 74-9
[5] Zhang N, Jia Z, Shi Z 2005 Text Classification based on KNN Algorithm *Computer Engineering*, **31**(8) 171-3
[6] Joachims T 1998 Text categorization with support vector machines: learning with many relevant features *Proceeding of ECML – 98 10th European Conference on Machine Learning* Berlin: Springer-Verlag 137-42
[7] Gora G and Wojna A 2002 A classifier combining rule induction and k-NN method with automated selection of optimal neighbourhood

Proceedings of the Thirteenth European Conference on Machine Learning Heidelberg Springer Berlin **2430** 111-23

[8] Dudai S A 1976 The distance-weighted k-nearest neighbour rule. *IEEE Transactions on Systems*, *Man and Cybernetics* **6**(4) 325-7

[9] Ferri F and Vidal E 1992 Colour image segmentation and labelling through multiedit-condensing *Pattern Recognition Letters* **13**(8) 561-8

[10] Segata N, Blanzieri E, Delany S J, Cunningham P 2010 Noise reduction for instance-based learning with a local maximal margin approach *Journal of Intelligent Information Systems* **35**(2) 301-31

[11] Hart P E 1968 *IEEE Transactions on Information Theory* IT **14**(3) 515-6

[12] Toussaint G T, Bhattacharya B K, Poulsen R S 1985 The Application of Voronoi Diagrams to Nonparametric Decision Rules *Proceedings Computer Science and Statistics: 16th Symposium on the Interface* North-Holland Amsterdam 97-108

[13] Wilson D R, Martinez T R 2000 Reduction techniques for instance-based learning algorithms *Machine Learning* **38**(3) 257-86

[14] Wu Y Q, Ianakiev K, Govindaraju V 2002 Improved k-nearest neighbour classication *Pattern Recognition* **35**(10) 2311-8

[15] Fayed H A, Atiya A F 2009 *IEEE Transactions on Neural Networks* **20**(5) 890-6

[16] Huang D, Chow T W S 2006 Enhancing density-based data reduction using entropy *Neural Computation* **18**(2) 470-95

[17] Paredes R, Vidal E 2006 Learning prototypes and distances: a prototype reduction technique based on nearest neighbour error minimization *Pattern Recognition* **39**(2) 171-9

[18] Brighton H, Mellish C 2002 Advances in instance selection for instance-based learning algorithms *Data Mining and Knowledge Discovery* **6**(2) 153-72

[19] Tan S 2005) Neighbor-weighted k-nearest neighbour for unbalanced text corpus. *Expert Systems with Applications* **28**(4) 667-71

Authors

Caixian Chen, born in September, 1979, China

Current position, grades: researcher at Shandong University of Finance and Economics and Shandong Provincial Key Laboratory of Digital Media Technology, China.
University studies: Master's degree in computer science and technology at Wuhan University of Technology, China in 2005.
Scientific interests: natural language processing and information retrieval.

Huijian Han, born in December, 1971, China

Current position, grades: professor at Shandong University of Finance and Economics. Researcher at Shandong Provincial Key Laboratory of Digital Media Technology, China
University studies: PhD degree in computer science and technology from Shandong University, China in 2010.
Scientific interests: CG&CAGD and animation.

Zheng Liu, born in April, 1980, China

Current position, grades: associate professor at Shandong University of Finance and Economics. Researcher at Shandong Provincial Key Laboratory of Digital Media Technology, China.
University studies: PhD degree in computer science and technology from Shandong University, China in 2010.
Scientific interests: information retrieval and data mining.

Cost-Sensitive learning on classification

Qin Yang[1]*, Changyao Zhou[2]

[1]*School of Business, Sichuan Agricultural University, Dujiangyan 611830, Sichuan, China*

[2]*School of Resources and environment, Sichuan Agricultural University, Wenjiang 611130, Sichuan, China*

Abstract

Real-world predictive data mining (classification or regression) problems are often cost sensitive, meaning that different types of prediction errors are not equally costly. In this paper we propose a new algorithm for cost-sensitive classification in a multiple time series prediction problems. The fitness function of the genetic algorithm is the average cost of classification when using the decision tree, including both the costs of tests (features, measurements) and the costs of classification errors. The proposed model is evaluated in a real world application based on a network of satellite network map distributed in land spatial pattern evolution in Chengdu Plain. These satellite networks generate multiple time series data representing land spatial pattern. This study presents a new algorithm for cost-sensitive classification that deal with class imbalance using both recompiling and CSL. The method combines and compares several sampling methods with CSL using support vector machines (SVM). We build our cost-benefit model for the prediction process as a function of satellite network in a distributed land spatial and measured the optimum number of satellite network that will balance the expenses of the system with the prediction accuracy.

1 Introduction

Machine learning and data mining reply heavily on a large amount of data to build learning models and make predictions, and thus, the quality of data is ultimately important. To evaluate in a real world application based on a network of satellite network distributed in land spatial pattern evolution in a region. We use real-world data-sets in land spatial pattern evolution from 2008 to 2013. Cost/Benefit Analysis is a relatively simple and widely used technique for deciding whether to make a change [1]. Profit and costs drive the utility of every land resource use decision. As land resource use decision making, from strategic to operational planning, is based upon future realizations of the decision parameters [2]. The real trick to doing a cost benefit analysis well is making sure you include all the costs and all the benefits and properly quantify them. Cost-benefit analysis has already attracted much attention from the machine learning and data mining communities [3]. Thus, cost-sensitive learning algorithms should make use of only known values. Under cost-sensitive learning, we impute values of data, and the learning algorithms make use of values to minimize the total cost of tests and classifications.

In the literature of cost sensitive analysis for data mining applications, most emphasis is given to classification problems. Recent research and related publications show that cost sensitive analysis is not deeper analysed, modelled, and applied to prediction problems. Conversely, data mining methods for regression and time series analysis generally disregard economic utility and apply simple accuracy measures [2]. Only some theoretical approaches exist for specific data mining methods such as neural networks and support vector machines.

In this research, we propose a new approach and develop a methodology to apply the cost benefit analysis to real world prediction problems. Our application is prediction of land spatial pattern. Data for our research is collected from a patented distributed satellite network of land spatial pattern. In this paper, we have proposed a model for cost-benefit analysis in a multiple time series prediction problems. We have identified three distinct areas of the cost function and analysed the behaviour in each region. We have experimentally found out the threshold values corresponding to these regions. We presents a new algorithm for cost-sensitive classification that deal with class imbalance using both resembling and CSL. The method combines and compares several sampling methods with CSL using support vector machines (SVM). Results are compared and analysed. We have discussed how this methodology can be utilized in similar distributed systems.

Following a brief introduction to previous work in section 2. Section 3 introduces multiple time series prediction problems and its formalization for prediction of land spatial pattern based on daily satellite network data. We report our recent research efforts in introducing cost sensitive analysis for this real world prediction application in Section 4. Conclusions are given in Section 5.

Corresponding author e-mail: yangqin@sicau.edu.cn

2 Related work

In recent years data mining community has attempted incorporating cost benefit analysis into classification problems, where the "Cost" could be interpreted as misclassification cost, training cost, test cost, or others [5]. Among all different types of costs, the misclassification cost is the most popular one. In general, misclassification cost is described the cost of predicting that an example belongs to class *i* when in fact it belongs to class *j* [6]. Hollmen et al., and Elkan introduce a cost model that inland resource uses the specific properties of objects to be classified. Instead of a fixed misclassification cost matrix, they utilize a more general matrix of cost functions. These functions operate on the data to be classified and are re-calculated for each data point separately [7, 8].

Most of the existing cost sensitive classifiers assume that datasets are either noise free or noise in the datasets is less significant. However, real-world data is never perfect and suffer from noise that may impact models created from data. Zhu and Wu have addressed the problems of class noise, which means the errors introduced in the class labels. They have studied the noise impacts on cost sensitive learning and proposed a cost guided noise handling approach for effective learning [6]. The class imbalance problem has been recognized as a crucial problem in machine learning and data mining. Such a problem is encountered in a large number of ranges, and it can lead to poor performance of the learning method [9]. It has been indicated that cost-sensitive learning is a good solution to the class imbalance problems and Zhou and Liu have studied methods that address the class imbalance problems applied to cost-sensitive neural networks [3].

Similarly, for predictive data mining problems of regression and time series analysis the costs arising from invalid point prediction, costs of under prediction versus over prediction, etc. are also analysed in the literature. Crone et al. have analysed the efficiency of a linear asymmetric cost function in inventory management decisions, training multilayer perceptions to find a cost efficient stock-level for a set of seasonal time series directly from the data [2]. A similar work has been proposed by Christoffersen and Diebold by introducing a new technique for solving prediction problems under asymmetric loss using piecewise-linear approximations to the loss function [10].

Wang and Stockton have investigated how the constraints imposed by changing export market affect the identification of "cost estimating relationships" and investigated their relative benefits and limitations in terms of their effects on the overall cost model development process. Neural network architecture has been used and a series of experiments have been undertaken to select an appropriate network [11].

Cost estimation generally involves predicting labour, material, utilities or other costs over time given a small subset of factual data on "cost drivers." Alice has examined the use of regression and SVM models in terms of performance, stability and ease of use to build cost estimating relationships. The results show that SVM have performed well when dealing with data, which there is little prior knowledge about the cost estimating relationship to select for regression modelling. However, regression models have shown significant improvements in terms of accuracy in cases where an appropriate cost estimating function can be identified [12].

3 Prediction of total power production

This paper presents our research results in analysis of distributed land spatial pattern in a region, and development of prediction model based on data collected from multiple distributed satellite network related to specific land spatial pattern. These land spatial pattern operate data throughout the year continuously.

Each land spatial keeps record of real-time land spatial pattern, current land spatial output and real time changes and variations in land spatial usage and supply availability. These data are taken at specific time intervals and they vary from a day to a week or a month. The sampling frequency depends on the type of data that are collected at the land spatial.

The data for our analysis comprises readings of satellite network at 417 distributed land spatial pattern, and the cumulative variable that correspond to the total land spatial pattern of those 417 land spatial pattern. The data set that we are using is collected daily and we use a repository of three years from year 2008 to year 2013. It has 365 data for each land spatial in years 2011, 2012 and 2013. These historic data can be used to build and improve relative models used for short and long term land spatial pattern forecasting. They are currently used by land trading and marketing firms and federal regulatory agencies including Homeland Security.

Prediction systems have to obtain these data sets at a cost. Getting data from more land spatial pattern to make prediction means increase in expenses. Our goal is to predict the total land spatial pattern with reduced number of satellite network, where we can compromise between the expenses for collecting data and the quality of prediction accuracy. Based on the analysis of different multiple time series prediction methods we have selected SVM as the best candidates for our study in building cost sensitive prediction model [13-15]. Our approach shows that it is possible to measure the total land spatial pattern using reduced number of satellite network stations. The modelling results and general approach may be used in other systems to determine the required number of satellite network to be used for data collection.

3.1 DATA COLLECTION AND PRE-PROCESSING

The data for analysis is collected daily for the years from year 2008 to year 2013. Most of the machine learning methods including SVM requires that all data sets to be normalized. We use MATLAB to normalize data between

[-1,+1] for 417 land spatial pattern. In the raw data set there are some entries with 0 values. They correspond to days in which a satellite network does not record a value for a given day. In our problems, all those 0 values are taken as actually recorded values after proving with the authorities [16].

We have used year 2008 data as training data set and year 2012 and year 2013 data as testing data. Main goals of our research are a) To measure and compare the quality of different prediction methodologies, and b) To measure optimum number of satellite network, which will give the best results balancing expenses of the system. We used several sampling methods with CSL using support vector machines (SVM). Accuracy of this model is measured, by computing the correlation coefficient, between actual values and predicted values in the testing data set. Prediction accuracy is lie on the amount of satellite network we use, i.e. number of power land spatial inputs we use in the prediction. To prove this we have built training and testing data sets by varying the number of satellite network from which the data are collected. In our study we considered data from 16, 64, 128, 512 and 1024 land spatial inputs. An overview of these steps is presented in Table 1. In the following sections, we elaborate on these steps.

TABLE 1 Classify land resources

Step: Assessment of land resources:
1. Land resources classification:
–identify the land resources role in the organization
2. Define land resources policy goals:
–assign weights to land resources policy goals (C,I,A)
3. Classify land resources:
–enumerate resources available on the given system
–determine the resource importance for each land resources policy goal
–compute the overall resource weight for each land resources policy goal

3.2 EXPERIMENTAL RESULTS

In SVM, a selected number of satellite network data (time series) are used in prediction. We have 417 satellite networks as data sources, and there are various ways of selecting the subset. Selection of a subset of satellite network (in our case 10, 20, 30, 40 tc.) will cause combinatorial explosion. We made one heuristic approximation in this step to eliminate computational complexity. Each of 417 satellite network time series is compared with the total land spatial pattern by computing the correlation factor. We wanted to measure how much each of satellite network measurements are correlated to the output. Then, we sorted satellite network based on the correlation factor, and selected subset is a portion of top ranked satellite network.

We used a feed forward neural network with back propagation learning using only one hidden layer. The algorithm was implemented in MATLAB ver6.5. Experiments showed that equal number of input and hidden nodes give the best results in prediction. Inputs to the network are the data columns corresponding to satellite network' recordings at each land spatial throughout the

year and the single output represents predicted value for the land spatial pattern in the region. We have experimented the SVM model with different combinations of the parameters and determined that the values 0.001 for accuracy parameter and 0.04 for learning rate with Tangent-Sigmoid activation function give the best prediction results. Figure 1 shows the results for prediction using the SVM with 40, 70, and 100 satellite network with the correlation coefficients (r) 0.84, 0.93, and 0.95 respectively. A toolbox called LIBSVM (A Library for Support Vector Machines) was used for SVM methodology [17].
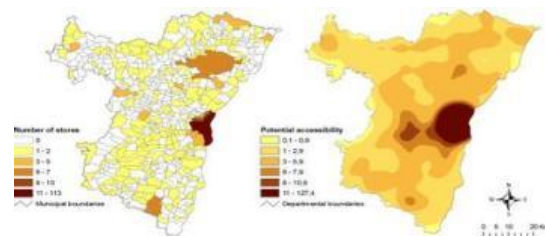
FIGURE 1 Graph for solution of all hyper dataset.

We use support vector machines (SVM) as the base classifier. Grid search is used to determine the best hyper parameters for SVM and the resembling techniques.

(a) Method 1: Combination of Sampling techniques with CSL, called S-CSL.

Algorithm 1: S-CSL
Begin
    while NFC < MAX$_{NFC}$ dob
        for i=1 to ps do
            Update velocity of the ith location point according to (1);
            Update position of the ith location point according to (2);
            Calculate the fitness value of the ith location point;
            NFC++;
        end for
        Update the pbest and gbest, if needed.
        for i=1 to ps do
            if rand(0,1) < p$_o$ then
                Generate a mutant location point according to (3);
                Calculate the fitness values of the mutant location point;
                NFC++;
                Select the fitter one between the current location point and the mutant location point as the new current location point;
            end if
        end for
        Update the pbest and gbest, if needed.
    end while
End

(b) Method 2: Using CSL by Optimizing Cost Ratio Locally, called CSL-OCRL.

Algorithm 2: CSL-OCRL
Begin
    while NFC < MAX$_{NFC}$ do
        for i=1 to ps do
Update velocity of the ith location point according to (1);
Update position of the ith location point according to (2);
            Calculate the fitness value of the ith location point;
            NFC++;
        end for
        Update the pbest and gbest, if needed;
        Calculate the boundaries of current search space;
        Create a random location point Y according to (3);
        Create a trail location point X* according to (5);

Calculate the fitness value of X*;
NFC++;
If X* is better than X_w then
    Replace X_w with X*;
end if
Update the pbest and gbest, if needed.
end while
End

Standard facilities of matrix algebra in MATLAB are used for MR analysis. The Figure 1 shows the variations of quality of prediction results obtained with SVM 128 input satellite network, r = 100

(c) Using SVM with 16 input satellite network, r = 10

(d) Using SVM with 1024 input satellite network, r =510

Input satellite network for three methods MR for different number of input time series.

Detailed comparison of these methods and discussion about optimal number of satellite network for different models are given in our previous article. The hypothesis that optimal number of satellite network should be the trade off between accuracy of prediction and costs of the monitoring system is the initiating point of the current research [16].

4 Cost benefit analysis in multiple time series prediction

Should we collect data from all areas? If not what would be the optimum number of areas to use for prediction? What are economic benefits from the prediction system? These questions are essential part in analysis of a solution for our prediction system based on distributed satellite network system which monitors land spatial pattern. A cost benefit analysis for multiple time series prediction, we are developing and applying in this research, will give some of the answers.

$$V(t+1) = wV(t) + c_1 r_1 (pbest_i(t) - X_i) +$$
$$c_2 r_2 (gbest_i(t) - X_i(t)), \tag{1}$$

$$X_i(t+1) = X_i(t) + V_i(t+1), \tag{2}$$

$$X_i^* = gbest = a_1(gbest - X_i) + (1 - a_1)(X_{i1} - X_{i2}). \tag{3}$$

In order to verify the performance of the proposed approach, ten famous benchmark functions are selected in our experiments [12]. According to their properties, they are divided into two types: functions f_1 to f_7 are uni-modal functions and f_8 to f_{10} are multimodal functions. All the functions are minimized problems. The specific definitions, dimensions, and the global optimum are listed as follows:

$$f_1 = \sum_{i=1}^{D} x_i^2 ,$$

where $x_i \in [-100, 100]$, D=30, and the global optimum is 0.

$$f_2 = \sum_{i=1}^{D} |x_i| + \prod_{i=1}^{D} x_i ,$$

where $x_i \in [-10, 10]$, D=30, and the global optimum is 0.

$$f_3 = \sum_{i=1}^{D} \left(\sum_{j=1}^{i} x_j \right)^2 ,$$

where $x_i \in [-100, 100]$, D=30, and the global optimum is 0.

$$f_4 = \max_i (|x_i|, 1 \le i \le D),$$

where $x_i \in [-100, 100]$, D=30, and the global optimum is 0.

$$f_5 = \sum_{i=1}^{D-1} \left[100 \left(x_{i+1} - x_i^2 \right)^2 + \left(x_i - 1 \right)^2 \right],$$

where $x_i \in [-30, 30]$, D=30, and the global optimum is 0.

$$f_6 = \sum_{i=1}^{D} \left(\lfloor x_i + 0.5 \rfloor \right)^2 ,$$

where $x_i \in [-100, 100]$, D=30, and the global optimum is 0.

$$f_7 = \sum_{i=1}^{D} i x_i^4 + rand[0,1),$$

where $x_i \in [-1.28, 1.28]$, D=30, and the global optimum is 0.

$$f_8 = \sum_{i=1}^{D} -x_i \sin \left(\sqrt{|x_i|} \right),$$

where $x_i \in [-500, 500]$, D=30, and the global optimum is -12569.5.

$$f_9 = \sum_{i=1}^{D} \left[x_i^2 - 10 \cos(2\pi x_i) + 10 \right],$$

where $x_i \in [-5.12, 5.12]$, D=30, and the global optimum is 0.

$$f_{10} = -20 \exp \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \right) -$$
$$\exp \left(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi x_i) \right) + 20 + e,$$

where $x_i \in [-32, 32]$, D=30, and the global optimum is 0.

There are several costs involved in building our prediction system. They can be of two types: fixed cost and variable cost. In optimization problems, fixed cost only shifts the cost curve into a higher or lower level. So, we only considered variable cost in our analysis. In building and maintaining the satellite network system, these expenses correspond to hardware installation and maintenance, data collection, data processing and data analysis.

These expenses would include, procedure OPTIMIZECOSTRATIO(DTrain;_; _)

```
Input: DTrain, SVM parameters _, step length _
Outputs: the best cost ratio for GMean
2: (DLocalTrain;DHoldout)  DTrain . split for 5-fold CV
3: ImbaRatio  jMajorj
jMinorj . imbalance ratio of DTrain
4: maxRatio  ImbaRatio _ 1:5
5: currentRatio  1:0
6: bestGMean  0
7: while currentRatio <= maxRatio do
8: buildLocalModel(DLocalTrain; _)
9: currentGMean  testLocalModel(DHoldout)
10: if (currentGMean > bestGMean) then
11: bestGMean  currentGMean
12: bestCostRatio  currentRatio
13: end if
14: currentRatio  currentRatio + _
15: end while
16: return bestCostRatio
17: end procedure
```

Cost may also include a risk factors that would compensate increase of prices, depreciation etc. For these reasons it should be determined very carefully by experts in the range.

Accuracy of prediction is a nonlinear function of the number of satellite network [16]. We selected the polynomial model with relative small error and at the same time enough simple. We experimentally confirmed that polynomial function of third order makes a good approximation of the prediction non-linearity:

```
1: procedure HYPERSEARCH(DTrain;)
returns the best hyperparameters _ for eval. metric E
2: (DLocalTrain;DHoldout)  DTrain //split for 5-fold CV
//Raw search:
3: bestC; besty ←0
4: for i ←15,……10 do
5: for j ←15,……0 do
6:   y←2j ; C←2i
7:     buildLocalSVM(DLocalTrain; ;C)
8:     TestLocalModel(DHoldout) //using metric E
9:       Update bestC; best
10: end for
11: end for //Smooth search:
12:  for i← bestC-1,…… bestC + 1; step r do
13:  for j← best -0.1,…… best + 0:1; step q do
14:   y←j; C←i
15:   buildLocalSVM(DLocalTrain; ;C)
16:   TestLocalModel(DHoldout) //using metric E
17:  _  C;  //Update the best parameter values
18:    end for
19:  end for
20   : return _
21: end procedure
```
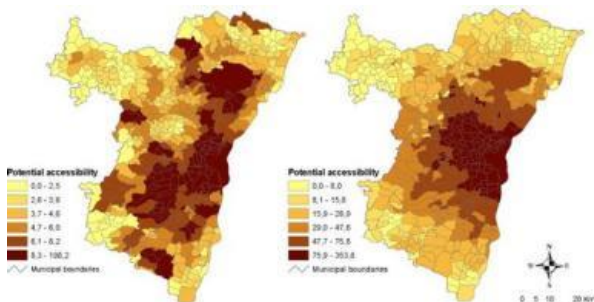

FIGURE 2 Graph for solution of hypothyroid dataset
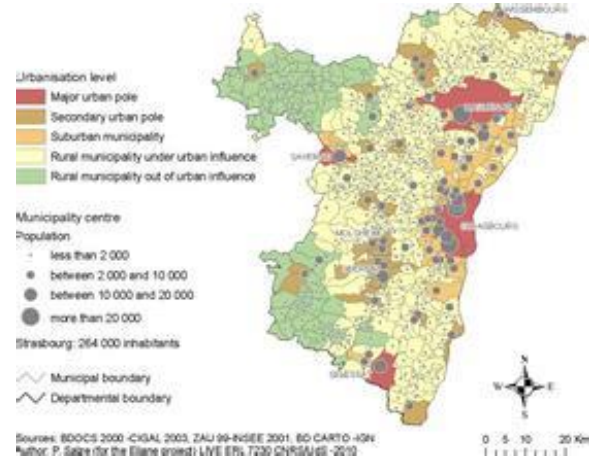
FIGURE 3 Graphical interpretation of Ann dataset

4.1 COST BENEFIT MODEL

We experimentally determined parameters in relative formulae $y(n)$ for the SVM (SVM y) methodologies. We can find the relative relationship between the accuracy and the number of inputs for each of the methods applying a curve fitting procedure and assuming that the function is a polynomial third order. The approximated equations fits the graphs of Figure 2:

These three functions represent our cost-benefit models for prediction. Each model is based on two input parameters:' C -ratio of cost constants, and n - number of satellite network.

4.2 INTERPRETATION OF EXPERIMENTAL RESULTS

We can analyse how the actual cost function changes with values of constant 1 2 C C or ' C, and we will try to find a minimum of these functions for different ' C values using experimental results.

4.2.1 Large benefits

Figure 3 gives a plot for cost function C for very small value of ' C 0.0001based on our experimental results. Three graphs represent the three applied prediction methodologies. Small ' C means that the value of 2 C is very high, relative to the value of 1 C i.e. the users measure that the benefit of prediction accuracy is very high and its overweight any cost 1 C for satellite network installation and maintenance. As it is expected, the cost function C is continuously decreasing function, where its minimum is with maximum number of satellite network n (in our case 417). That means, we accept all expenses for installation of all 417 satellite network and there is no prediction system. Output will be just calculated as a sum of all satellite network values. Even if the satellite network are expensive, it is not a sufficient reason to influence the cost function, which is minimum for maximum *n*.

4.2.2 Large costs

Figure 3 shows the variation of total cost as a function of a number of input satellite networks for cost 64. It gives its minimum value with minimum number of input areas close to 0. If the installation of satellite network determined by cost CSL is very costly, and benefits (profit) determined by Cost CSL is relatively low, the function will be continuously increasing. The Figure 3. Graph for solution may found at extremely small number of satellite network (close to 0). Again, we do not need prediction system because the expenses are so high, that the system is not economically feasible.

4.2.3 Costs and benefits are balanced

The most interesting case is when the cost ratio is increasing. We believe that it is the common case in real world applications. Minimum value for total cost , will determine necessary number of satellite network to obtain maximum benefits from the prediction system. Based on the number of satellite network we can measure the quality of prediction for the recommended configuration.

It shows the case where the minimum of cost, occurs between the minimum and the maximum number of areas. As expected, simple SVM model shows the minimum cost where the required number of input satellite network is around 128. Both simple SVM and CSL using support vector machines (SVM). It shows a minimum at around 100 input satellite networks for this value. As shows the variation the total cost with the different values, and the number of satellite network. In addition, data shows the accuracy of prediction at the optimum number of satellite network for each case. As we discussed so far, it is clear that the value plays a major role in selecting the minimum

value for the cost function. In general cost will determine, if it is useful to build the prediction system or not. Based on the experimental results, we can see that the behaviour of the cost function defines three regions. The cost function behaves as an asymptotically increasing function, where the minimum cost can be found at minimum number of satellite network in all three methods. Accuracy at this situation is at its minimum values. In both previous cases we do not need a prediction system. The cost function clearly defines a minimum value for the number of satellite network at a point inside the range.

Accuracy and the number of satellite network used depend on prediction technique. This is a good engineering design heuristics, obtained from cost benefit analysis, to determine whether prediction analysis is economically feasible for our system.

5 Conclusions

In this paper we describe a new methodology for the cost-benefit analysis in multiple time series prediction. It is applied in a real world distributed satellite network. Satellite network generate time series data representing land spatial pattern evolution in Chengdu Plain. For establishing the prediction model, we use three common prediction methods SVM. Training and testing data were available for year's 2008 to 2013 land spatial pattern evolution in Chengdu Plain. Our results show that CSL using support vector machines (SVM). Technique gives the best prediction model.

References

[1] Li P, Qiao P L, Liu Y C 2008 A hybrid re-sampling method for SVM learning from imbalanced data sets *Proceedings of fifth Int. Conf. on Fuzzy Systems and Knowledge Discovery* Washington DC USA 65-9

[2] Zhang Y, Zhou, Z H 2008 Cost-sensitive face recognition *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Anchorage AK 416-422

[3] Yang Q, Wu X 2006 10 challenging problems in data mining research *International Journal of Information Technology and Decision Making* **5**(4) 597-604

[4] Chawla N V, Bowyer K, Hall L, Kegelmeyer W P 2002 SMOTE: synthetic minority over-sampling technique *Journal of AI Research* **16** 321-57

[5] Chawla N V, Japkowicz N Kotcz A 2004 Special issue on learning from imbalanced data sets *SIGKDD Explorations* **6** 1-6

[6] Nickerson A, Japkowicz N, Millos E 2001 Using unsupervised learning to guide resampling in imbalanced data sets *The Eighth International Workshop on AI and Statitsics* 261-5

[7] Raskutti B, Kowalczyk A 2004 Extreme re-balancing for SVMs: acase study *SIGKDD Explorations* **6**(1) 60-9

[8] Chen X, Gerlach B, Casasent D 2005 Pruning support vectors for imbalanced data classification *IEEE International Joint Conference on Neural Networks* **3** 1883-8

[9] Lessmann S 2004 Solving imbalanced classification problems with support vector machines *Proceedings of International Conference on*

Artificial Intelligence 214-20

[10] Liu X Y, Wu J, Zhou Z H 2009 Exploratory under-sampling for class-imbalance learning *IEEE Transactions on Systems, Man and Cyberernetics Part B* 539-50

[11] Li X, Wang L, Sung E 2008 AdaBoost with SVM-based component classifiers *Engineering Applications of Artificial Intelligence* **21**(5) 785-95

[12] Sheng S, Ling C X, Ni A, Zhang S 2006 Cost-sensitive test strategies *The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference* **1** 323-31

[13] Ting K M 2002 A study on the effect of class distribution using costsensitive learning *Discovery Science* 98-112

[14] Akbani R, Kwek S, Japkowicz N 2004 Applying support vector machines to imbalanced datasets *Proceedings of the 15th ECML* 2004 39-50

[15] Batista G E A P A, Prati R C, Monard M C 2004 A study of the behavior of several methods for balancing machine learning training Data *SIGKDD Explorations* **6**(1) 20-6

[16] Van Hulse J, Khoshgoftaar T M, Napolitano A 2007 Experimental perspectives on learning from imbalanced data *The 24th ICML* ACM 935-42

[17] Platt J C 1999 Probabilistic outputs for SVM and comparisons to regularized likelihood methods A*dvances in Large Margin Classifiers* MIT Press 61-74

[18] Domingos P 1999 Metacost: A general method for making classifiers costsensitive *The 5th International conference on Knowledge Discovery and Data mining KDD1999* 155-164

[19] Margineantu D D 2000 When does imbalanced data require more than cost-sensitive learning? *Papers from the AAAI Workshop Technical Report WS-00-05* 47-50

Authors

Qin Yang, born in March, 1977, China

Current position, grades: researcher at Sichuan Agricultural University, China.
University studies: MS degree in Computer Software and Theory, University of Electronic Science and Technology of China.
Scientific interests: Computational intelligence: location point swarm optimization, ant colony optimization, genetic algorithm, differential evolution.

Changyao Zhou, born in March, 1989, China

Current position, grades: Master Student in Sichuan Agricultural University, China.
Scientific interests: Computational intelligence: location point swarm optimization, sustainable use of land resource.

A medical quality evaluation method based on combined weight

Shenyi Qian*, Yanling Zhu, Shen Li

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China

Abstract

Medical quality evaluation is the key and important link of the current medical institutions improve the core competitiveness, considering the characteristics of the medical industry, the paper constructs the surgical and non-surgical medical quality evaluation index system. In addition, the traditional medical quality evaluation in determining the index weight coefficients are too single, it is easy to cause the subjective assessment results too much or the accuracy is not high. Therefore, the paper using the Analytic Hierarchy Process (AHP) for subjective weight and using the Rough Set (RS), then get them together and put forward an approach of medical quality evaluation method based on combined weight, this method absorbs the advantage of them and overcomes the disadvantages of them and achieve the complementary advantages. Finally through the case analysis, verifying the feasibility and effectiveness of the method.

Keywords: medical quality evaluation, indicator system, combined weight, AHP, RS

1 Introduction

With the development of medical and health services, the evaluation of medical quality has become one of the key of medical institutions improve the core competitiveness, the evaluation of medical quality not only can evaluate the medical quality of the medical institutions, also can objectively reflect the existing problems and the weak links in the medical institutions, then provide decisions for the managers of the medical institutions and further to help the medical institutions to change and improve the medical quality in some specific ways. Domestic experts in the field of medical management has carried on the exploration and research for a long time. According to different provinces and different levels of medical institutions, the corresponding evaluation indicator system is also different each other, but is gradually perfect. The medical management department of the centre and provinces has already taken all kinds of means and measures to evaluate the medical quality of the medical institutions. The current evaluation of medical quality mainly includes qualitative analysis and quantitative analysis. Qualitative analysis includes expert experience, scores of patient satisfaction and hospital grading system, etc. Only use qualitative analysis often cause subjective evaluation results. Single use qualitative analysis often cause subjective evaluation results; Quantitative analysis includes calculate indicator data, analyse the number of changes, etc. although depend on quantitative analysis can get objective evaluation results, but if the evaluative data is not real also lead to the deviation of results. Therefore, at present the main problems of medical quality evaluation is how to combine the subjective evaluation results and objective evaluation results and get a comprehensive evaluation results.

This paper basis for the indicator system of medical quality evaluation, combine the subjective weight based on analytic hierarchy process (AHP) and the objective weight based on rough set, then build the optimization model of the comprehensive weights. Finally, though the function of Lagrange verify the feasibility of the model.

2 Medical quality indicator system

The indicator of medical quality evaluation is a scientific concept which reflects the pros and cons of medical quality, on the basis of the statistics, getting different indicator together and building the indicator system of medical quality evaluation are the premises and prerequisites of the medical quality evaluation. In this paper, the acquisition of the indicator system of medical quality evaluation and the construction of indicator system, mainly from the following several aspects:

1) Reference 2005-2010, the ministry of health department completes <the establishment and application research of indicator system of China hospital medical quality evaluation>, it propose the Chinese Medical Quality Indicator System (CHQIS), the system set up the three categories of 11 1 level indicators and 33 2 level indicators which include the in-hospital death related, non-plans to return relevant, related adverse events. Currently CHQIS has 730 single indicators and 4610 composite indicators.

2) Consult and reference many documents which elated to the medical service quality evaluation at home and abroad, foreign medical quality evaluation develop earlier and the choice of many indicators are through strict

---

Corresponding author e-mail: qsy@zzuli.edu.cn

Operation Research and Decision Making

screening and clinical trials that can reasonably reflect the medical quality, it has good scientific and sensitivity. Such as the International Quality Indicator Proiect [1, 2] (IQIP) that divided into 25 categories and 285 indicators, it used for evaluate all levels of hospital and medical institutions, etc. IQIP pays attention to the results of the medical service and patients interests; focus on the influence of "negative event"; stress on the comparability of indicators; the choice standard of indicators are more rigorous.

TABLE 1 Non-surgical medical service quality evaluation indicator system

Indicator classification	First indicator	Second indicator
Non-surgical	Therapeutic effect	Cure rate
		Improvement rate
		Not cured rate
		Death rate
	Work efficiency	Inpatient Amount
		3-day-correct-diagnosis rate
		Average length of stay
	Diagnostic level	The coincidence rate of Admission and Discharge
		The coincidence rate of clinic and pathology
		The coincidence rate of radiation and pathology
		Adverse reaction rate of blood transfusion
		Adverse reaction rate of transfusion
	Medical records writing	Medical record rate class a
		Medical record rate class b
		Medical record rate class c
	Cost	All-in cost

TABLE 2 Surgical medical service quality evaluation indicator system

Indicator classification	First indicator	Second indicator
Non-surgical	Therapeutic effect	Cure rate
		Improvement rate
		Not cured rate
		Death rate
		Healing rate class a
		Healing rate class b
		Healing rate class c
	Work efficiency	Inpatient Amount
		3-day-correct-diagnosis rate
		Average length of stay
		The coincidence rate of Admission and Discharge
		The coincidence rate of clinic and pathology
	Diagnostic level	The coincidence rate of radiation and pathology
		Adverse reaction rate of blood transfusion
		Adverse reaction rate of transfusion
		The coincidence rate of preoperative and postoperative
	Medical records writing	Medical record rate class a
		Medical record rate class b
		Medical record rate class c
	Cost	All-in cost

1) Quantitative analysis [3] based on medical data, while due to the large amount of data, data scattered, do not have a unified format, etc. Causing some data of indicators cannot acquire or the data quality cannot meet the conditions of evaluation, therefore in the process of the selection of indicators, we need to refer medical data, and then confirm the collection of the indicators.

2) Refer to some opinions and suggestions, which are given by relevant experts, these experts, are expert in the field of health care or medical management for a long time, they bear rich experience and provide authoritative evaluation indicators, these indicators can scientifically and reasonably reflect the medical quality.

Based on the research of the above several aspects and established the medical quality evaluation indicator system that direct at the surgical and non-surgical. Among them non-surgical categories including five primary indicator and 16 secondary indicator, surgery including 5 primary indicator and 20 secondary indicator (Table 1 and 2).

3 Confirm indicator weight

3.1 ANALYTIC HIERARCHY PROCESS

The Analytic Hierarchy Process (AHP) was first proposed by Thomas l. Saaty who is the famous American operations research expert and the professor at the university of Pittsburgh in the 1970 s [4]. AHP not only is a good subjective weighting method, but also a multi-criteria method of thinking. It makes the processes of people's thought hierarchical and quantitative [5, 6] and applies to the target complex and lacks the necessary data, meanwhile it brings us a problem-solving ideas that from the perspective of problem as a whole, by constructing a hierarchical structure, evaluating the influence of each part on the whole, so as to achieve the purpose of solving the problem.

The specific steps of calculate the subjective weight are as follows:

1) Establish hierarchical structure: the purpose of establishing hierarchical structure is to bring the complex issues organized and hierarchical, according to the target layer, criterion layer and measures layer construct the hierarchical model. The target layer only has one element, criterion layer can be divided into different levels and groups, the elements of different levels belong to the subordinate relation. Measures layer is located in the bottom of the hierarchy, the layer provides solutions to achieve objectives (elements) and has a plurality of measures.

2) Construct judgment matrix and assignment: after confirming the hierarchical structure, we need to compare between two indicator elements under the same layer that impact on the upper indicator elements at the criterion layer and construct judgment matrix, Thomas l. Saaty proposed the 1-9 scaling method to measure the importance between the two elements of two indicators, as shown in Table 3:

TABLE 3 Element 1-9 important degree

a_{ij}	Two indicators compared	Explain
1	Equal important	Indicator i and j equal important
3	Little important	Indicator i and j little important
5	More important	Indicator i and j more important
7	Obviously important	Indicator i and j obviously important
9	Absolutely important	Indicator i and j absolutely important
2, 4, 6, 8	Between two adjacent important degree	
The reciprocal of the above	Two goals in turn	

Judgment matrix $A = (a_{ij}) n \times n$, A meet the following properties: 1) $a_{ij} > 0$; 2) $a_{ij} = 1/a_{ji}$, $(i, j = 1, 2.....n)$; 3) $a_{ii} = 1$.

1) Hierarchical single sorting and inspection: after determine all the matrixes, we need to sort each matrix, namely solving matrix eigenvector, thus obtained weight value of each indicator. Taking an example of matrix $B = (b_{ij}) n \times n$, the steps of calculate the weight vector are as follow:

a) For each column as normalized $B_{ij} = \dfrac{b_{ij}}{\sum\limits_{i=1}^{n} b_{ij}}$,

$(i = 1, 2.....n)$;

b) For the sum of each row by row $W_i = \sum\limits_{i=1}^{n} B_{ij}$,

$(i, j = 1, 2.....n)$;

c) For vector $W_i = (W_1, W_2....W_n)^t$ as normalized, get $w_i = (w_1, w_2....w_n)^t$ as the approximate solution of the characteristic vector for matrix, namely the weight coefficient of indicator. After get the weights of indicators, we need check consistency of judgment matrix and ensure the matrix bear transitivity and consistency.

The steps to check consistency are as follow:
a) Calculate the maximum eigenvalue of the judgment matrix $\lambda_{\max} = \sum\limits_{i=1}^{n} \dfrac{(Bw)_i}{nw_i}$;

b) Calculate consistency indicator C.I.: $C.I. = \dfrac{\lambda_{\max} - n}{n - 1}$;

c) Look-up table to determine the corresponding average random consistency indicator R.I., according to different order of judgment matrix to check the table and get the average random consistency indicator R.I., the average random consistency indicator R.I. as shown in Table 4:

TABLE 4 Average random consistency indicator R.I.

Matrix order	1	2	3	4	5	6	7	8
R.I.	0	0	0.52	0.89	1.12	1.26	1.36	1.41

d) Calculate and determine the consistency of proportion C.R., $C.R. = \dfrac{C.I.}{R.I.}$. When $C.R. < 0.1$, the consistency of judgment matrix is acceptable, $C.R. > 0.1$, we consider the consistency of judgment matrix does not meet the requirements and need to re-amend the judgment matrix.

2) Hierarchical total sorting and inspection: In general, the sorting result of the last layer in the criterion layer, calculation and test steps are similar to the single sorting and inspection.

3.2 ROUGH SET THEORY

In 1982 Z. Pawlak formally proposed rough set theory [7], which is a new mathematical tool to analyse and deal with incomplete data and fuzzy knowledge representation. By using this incomplete and vague information, analysis and processing, we can find some hidden knowledge and useful information. Rough set theory has the advantage that in addition to data set, does not need require any preparative or additional information about the data, so the description of the uncertainty and processing are relatively objective [8].

Definition 1: A knowledge representation system $K = (U, Q, V, F)$, U is the domain, Q is a set of attributes, divided into condition attribute set C and decision attribute set D, $Q = C \bigcup D$, $C \bigcap D = \varnothing$, $V = \bigcup_{a \in A} V_a$ is a collection of property values, V_a represents a range of attributes $a \in Q$, f is $U \times A \rightarrow V$ mapping.

Definition 2: Given a domain U, $x, y \in U$, $P \subseteq Q$, if meet $\forall q \in P: f_q(x) = f_q(y)$, said object x and y to attribute set P is not identified. Denoted $Ind(P)$, that is the intersection of all equivalence relations. Expressed as $Ind(P) = \{(x, y) \in U \times U \mid \forall p \in P, p(x) = p(y)\}$.

Definition 3: An information system $S = (U, A)$, meet object set $X \subseteq U$ and attribute set $R \subseteq A$. Under U certainly belongs to the set of all the objects that consists of a set X is called X lower approximation, expressed as $R_{\_}(X) = \{x \in U \mid [x]_R \subseteq X\}$. Under U certainly or maybe belongs to the set of all the objects that consists of a set X is called X upper approximation, expressed as $R^{-}(X) = \{x \in U \mid [x]_R \cap X \neq \varnothing\}$. R is the positive domain of X $POS_R(X) = R_{\_}(X)$; R is the negative domain of X $NEG_R(X) = U - R^{-}(X)$. The boundary of X $BN_R(X) = R^{-}(X) - R_{\_}(X)$.

Definition 4: The information entropy $H(P)$ of knowledge P is defined as $H(P) = -\sum\limits_{i=1}^{n} p(X_i) \log(p(X_i))$.

Definition 5: the condition entropy $H(Q \mid P)$ of knowledge $Q(U / ind(Q)) = \{Y_1, Y_2, ...Y_m\}$ to knowledge $P(U / ind(P)) = \{X_1, X_2, ...X_m\}$ is defined as:

389

$$H(Q \mid P) = -\sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} p(Y_j \mid X_i) \log(p(Y_j \mid X_i)),$$

$$p(Y_j \mid X_i) = \frac{|Y_j \bigcap X_i|}{|X_i|}, i = 1, 2....n, j = 1, 2....m.$$

Definition 6: $S = (U, R, V, f)$ is a decision making system, $C \cup D = R$, C is condition attribute, D is decision attribute, $A \subset C$, the importance of any attribute $x_i \in C - A$, $SGF(x_i, A, D)$ is defined as:

$$SGF(x_i, A, D) = H(D \mid A) - H(D \mid A \bigcup \{x_i\}),$$

the greater the value of A, the greater the importance of the attribute, namely the indicator weight coefficient is larger, on the contrary, the smaller the importance of attribute, the smaller the indicator weight coefficient.

3.3 COMBINATION WEIGHTING

Assuming a decision information system $K = (U, Q, V, F)$, the subjective weight coefficient $w_{ai} = (w_{a1}, w_{a2}....w_{an})^t$ $(i = 1, 2.....n)$ obtained by the AHP, the objective weight coefficient $w_{bj} = (w_{b1}, w_{b2}....w_{bn})^t$ $(j = 1, 2.....n)$ obtained by the RS, w_{ck} is the combination of both weight coefficient, w_{ai}, w_{bj}, w_{ck} meet the following conditions:

$$\begin{cases} \sum_{i=1}^{m} w_{ai} = \sum_{j=1}^{m} w_{bj} = \sum_{k=1}^{m} w_{ck} = 1, (i, j, k = 1, 2.....n) \\ 0 \le w_{ai} \le 1 \\ 0 \le w_{bj} \le 1 \\ 0 \le w_{ck} \le 1 \end{cases}.$$

Establish an optimization model in the feasible region Ω:

$$\min \left\{ \sum_{i,j,k=1}^{m} [\mu(w_{ck}^2 / 2 + w_{ai}^2 / 2 - w_{ck} * w_{ai}) + (1 - \mu)(w_{ck}^2 / 2 + w_{bj}^2 / 2 - w_{ck} * w_{bj})] \right\},$$

μ is experience factor, $0 \le \mu \le 1$, the feasible region Ω meet $\Omega = \{w_{ck} \mid \sum_{k=1}^{m} w_{ck} = 1, 0 \le w_{ck} \le 1, (k = 1, 2.....n)\}$ and the optimization model has only one solution $w_{ck} = \mu w_{ai} + (1 - \mu) w_{bj}$, $(i, j, k = 1, 2.....n)$.

Proof: for the Lagrange function:

$$L(w_{ck}, \lambda) = \sum_{i,j,k=1}^{m} [\mu(w_{ck}^2 / 2 + w_{ai}^2 / 2 - w_{ck} * w_{ai}) + (1 - \mu)(w_{ck}^2 / 2 + w_{bj}^2 / 2 - w_{ck} * w_{bj})] + \lambda \left(\left| \sum_{k=1}^{m} w_{ck} - 1 \right| \right),$$

$$\frac{\delta L}{\delta w_{ck}} = 0, \sum_{k=1}^{m} w_{ck} - 1 = 0,$$

solve equations:

$$\begin{cases} \sum_{k=1}^{m} w_{ck} - 1 = 0 \\ \mu(w_{ck} - w_{ai}) + (1 - \mu)(w_{ck} - w_{bj}) = 0 \end{cases},$$

get $w_{ck} = \mu w_{ai} + (1 - \mu) w_{bj}$, $(i, j, k = 1, 2.....n)$.

According to the above proof we can know that the combination empowerment not only overcomes the subjective factors of excessive reliance on experience, but also avoid the only rely on the objective factors of data, thus improve the accuracy of the evaluation results.

4 Example analysis

In order to verify the feasibility of the evaluation method, based on the HIS data in a hospital of Zhengzhou City, using the combined weight method to evaluate five doctors' medical quality in the diagnosis of coronary heart disease (non-surgical). Due to the indicator system is numerous, so we select cure rate (A), the coincidence rate of admission and discharge (B), medical record rate class a (C), average length of stay (D), all-in cost (E) as the evaluation indicators.

4.1 AHP CALCULATE SUBJECTIVE WEIGHT COEFFICIENT

First of all, establish hierarchical structure, $D_i (i = 1, 2, 3, 4, 5)$ represent five doctors shown in Figure 1.
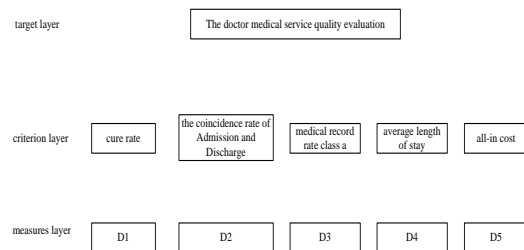
FIGURE 1 Hierarchical structure.

Then construct judgment matrix, as shown in Table 5:

TABLE 5 Judgment matrix

	A	**B**	**C**	**D**	**E**
A	1	5	7	4	3
B	1/5	1	3	1/4	1/5
C	1/7	1/3	1	1/5	1/6
D	1/4	4	5	1	1/3
E	1/3	5	6	3	1

Then calculate the weight vector, get the weight coefficient of indicator w_a = (0.4699, 0.0733, 0.0484, 0.1448, 0.2636). Finally check the consistency, when $n = 5$, $R.I. = 1.12$, the maximum eigenvalue $\lambda_{max} = 5.2465$,

$C.R. = 0.055 < 0.1$, the matrix meets the requirement. We can be seen that the weight coefficient accounted for the largest, followed by the all-in cost, average length of stay, the coincidence rate of admission and discharge, and finally the smallest proportion of medical record rate class a.

4.2 RS CALCULATE OBJIECTIVE WEIGHT COEFFICENT

Firstly, normalize and discretize the initial sample data (Table 6), obtain evaluation data set and construct decision table (Table 7), regard the five evaluation indicators as the condition attributes, reference w_a as decision attribute (F).

TABLE 6 The initial sample data table

	A	B	C	D	E
D1	97.4%	96.1%	98.7%	26	2345
D2	90%	98%	100%	23	2468
D3	91.7%	93.3%	98.3%	5	2180
D4	78.8%	100%	92.9%	29	3216
D5	69.2%	84.6%	100%	35	3238

TABLE 7 Decision table

	A	B	C	D	E	F
D1	1	2	2	3	1	1
D2	1	1	1	2	2	2
D3	1	2	2	1	1	1
D4	2	1	3	3	3	3
D5	3	3	1	3	3	3

According to the definition by using the information entropy theory, solve the importance degree of condition attributes, the weight coefficient of indicator.

4.3 COMBINATION WEIGHT CALCULATE THE WEIGHT

After confirm the subjective weight coefficient w_a and objective weight coefficient w_b, generating into the

Equation $w_{ck} = \mu w_{ai} + (1-\mu)w_{bj}$, $(i,j,k=1,2,...,n)$. When tend to subjective experience, $\mu \in [0.5,1]$; when tend to objective experience, $\mu \in [0,0.5]$; here take $\mu = 0.38$, make the subjective and objective weight coefficient ratio of the golden number. Calculated combination weight coefficient and the subjective and objective weight as shown in Table 8:

TABLE 8 Combination weight coefficient

	A	B	C	D	E
AHP	0.4699	0.0733	0.0484	0.1448	0.2636
Rough	0.2573	0.1394	0.1333	0.2444	0.2254
Combination weight	0.3381	0.1143	0.1010	0.2067	0.2399

As can be seen from the results, the combination weight and the subjective and objective weight basically have the same sorting, weight coefficient between the two weights, it make the result more accurate and reasonable. Finally, the combination of combination weight coefficient with sample data, to get the final rank of five doctors, D3 ranked the highest, followed by other D1, D2, D4, D5.

5 Conclusions

Against the disadvantages of the AHP and RS, this paper combines the two methods and discusses the application of combination weight in the medical quality evaluation. Through the example analysis, it proved that the method is feasible and improve the accuracy of the evaluation result to a certain extent and provide a new method for medical evaluation and research direction.

References

[1] Thomson R, Taber S, Lally J, Kazandjian V 2004 UK Quality Indicator Project (UK QIP) and the UK independent healthcare sector: a new development *International Journal for Quality in Health Care* **16**(Supp l) 1-6

[2] Cantrill J A, Sibbald B, Buetow S 1998 Indicators of the appropriateness of long term prescribing in general practice in the United Kingdom: Consensus development, face and content validity, feasibility and reliability *Quality in Health Care* **7**(3) 130-5

[3] Meirovich G, Brender-Ilan Y, Meirovich A 2007 Quality of hospital service: the impact of formalization and decentralization *International Journal of Health Care Quality Assurance* **20**(2-3) 240-52

[4] Saaty A L 1980 The analytic hierarchy process *McGraw Hill New York*

[5] Chin K S, Xu D L, Yang J B, Lam J P K 2008 Group-based ER-AHP system for product project screening *Expert Systems with Applications* **35**(4) 1909-29

[6] Wang Y M, Chin K S 2010 Fuzzy analytic hierarchy process: A logarithmic fuzzy preference programming methodology *International Journal of Approximate Reasoning* **52**(4) 541-53

[7] Pawlak Z 1982 Rough Sets *International Journal of Information and Computer Science* **11**(5) 34l-56

[8] Ali M I 2011 A note on soft sets, rough soft sets and fuzzy soft sets *Applied Soft Computing Journal* **11**(4) 3329-32

Authors

Shenyi Qian, born in November, 1975, China

Current position, grades: associate professor at Zhengzhou University of Light Industry.
University studies: Master degree in computer technology at Huazhong University of Science and Technology, China in 2001.
Scientific interests: information integration, database and business intelligence.

Yanling Zhu, born in January, 1984, China

University studies: Bachelor of Science degree from Zhengzhou University of Light Industry.
Scientific interests: data mining, business intelligence, computer software and theory.

Shen Li, born in October, 1988, China

Current position, grades: teaching assistant at HP international software talents and technology center.
University studies: master of engineering in computer technology from Zhengzhou University of Light industry.
Scientific interests: database and information integration.

Research of body dimension of Chinese adult male and application

Xuejing Du*, Huanhuan Guo, Zhanyu Wang

Institute of Transportation, University of Northeast Forestry, 150040, Harbin, China

Abstract

The goal of this study is to obtain accurate human dimension of current Chinese shape. Anthropometric data of adult male in GB10000-88 is adopted as sample observation data in this paper. Based on similarity theory of human body and the correlation of human body static dimensions, by statistical analysis of the sample date, the regression equation is established which determines the relationship between measurement of each part of the male body and height, weight. The maximum measurement error is within 5% by comparing the measurement data with calculating dimension data obtained by regression equations established in the paper, which verifies the rationality and accuracy of the regression equations. Once parameters of height and weight are given, the regression equations can quickly and easily provide effective body dimensions data used to build crash simulation model specifically for Chinese and man-machine product.

Keywords: automobile crash, dummy model, human dimensions, occupant restraint, regression equation

1 Introduction

Human body dimension determines the geometry space and range the human body occupied, which is a basic data for industrial product design, architectural design, and military industry, the technical reformation, equipment update and labour safety and protection [1]. The crash test dummy model made according to accurate human basic data is beneficial to design and evaluate automobile occupant restraint system. Chinese had carried on the human body measurements for the first time from1986 to1988 according to standards GB3975 [2] and GB5703 [3], and established a national database of human body dimension. Then, the national standard GB 10000-88 is drew up [4]. However, the data of human body dimension has strong timeliness. With the living standards improving greatly in China, body dimensions changes correspondingly since 1988. Hence, many body data can't meet the needs of life and survival design. Accurate information of Chinese human body size is important.

In order to obtain the accurate information of human body, Yu et al [5] studied body surface area measurement, using 3D scanner, and derived a simple body surface area measurement estimation formula for Chinese adults. Han et al [6] analysed the difference between body scan measurement and manual measurement, and the comparative analysis provided guidelines to use body scan measurement for obtaining accurate body measurement result. Wang used three statistical methods [7] to collect and analyse the body measurement data in static and dynamic measurement. In China, the China National Institute of Standardization carried out a sampling measurement pilot investigation about Chinese adult body dimension in Beijing and other three big cities in china in 2009, the collecting data showed that the adult male's height and waist circumference have increased 2 centimetres and 5 centimetres, and the size variance of bust and waist has diminished. On November 27, 2013, the project of "Chinese adults ergonomics basic parameter survey" was launched officially by China National Institute of Standardization as a leader, and was expected to complete by the end of May 2018, then the database of Chinese human dimensions would been fully updated at that time. In this case that the database has not been updated yet now, Bai and Cao [8] made the height and weight data of Chinese human acquired from the National physique monitoring reports in 2000 as two basic parameters, and then used empirical formulas of human body to obtain the main parts sizes and posture dimensions of Chinese human, and modified the frontal crash dummy to meet Chinese human characteristics. Zhang [9] divided the human body model into 15 parts, and then made correlation analysis and regression analysis of the 15 parts about height and weight respectively.

Zhang [10] proposed and proved mathematically the similarity theory of human body; he found that the same group human has similar shape. According to the similarity theory of human body [10] and the correlation of human body static dimensions, the relationship between human body parts dimension and human height and weight of Chinese adults is studied based on the data of human statistical parameter provided by GB10000-88. Then, the prediction linear regression equations of measurement parts of human adults male is established by using methods of correlation analysis and regression analysis. According to the linear regression equations, Chinese human

dimensions with body height and weight can be predicted, which can provide current body data for human body simulation modelling and guide man-machine product design. The simulation dummy of frontal crash is made in Madymo environment, which is as an application of the updated Chinese human dimensions in this study.

2 Methodologies

In order to study Chinese current human body dimensions, the sample observation data of body measurement in static state is from national database established in 1988. After analysis and discussion, height and weight are as two regression variables, and regression analysis method is applied to build the binary linear regression equation of human dimension. Human dimensions can be derived in single body and group human according to the regression equation in this paper.

Regression analysis is a statistical analysis method which studies the correlation relationship between a random variable and a (or several) controlled variable. The forecast model of human body dimensions is established based on correlation theory and linear regression analysis methods, which can better reflect the variability of each dimension of human body part and make the results more reasonable and credible. Regression analysis mainly includes three aspects:

1) Establish the empirical formula containing related variable (mathematical relationship).

2) Significance test:

Firstly, test whether the empirical formula established is valid, and then give the test of significance of each variable factor respectively;

Secondly, modify the empirical formula according to the situation;

Finally, determine the optimal regression equation.

3) Prediction and control according to the empirical formula.

2.1 ESTABLISH THE EMPRICAL FORMULA

After analysis, height and weight of human body are selected as two variable factors to establish the binary linear regression equations of Chinese human adults body dimensions, the mathematical relational expression is:

$$y = B_0 + B_1 x_1 + B_1 x_2, \qquad (1)$$

where, y is the dependent variable, dimension of part of human body that changing with height or weight; x_1, x_2 are variables of height and weight; B_0 is regression constant; B_1, B_2 are regression coefficients of height and weight.

When the regression empirical formula is determined, the next step is to analyse and calculate the sample data, and then work out the unknown regression coefficients B_0, B_1 and B_2 in Equation (1). In view of the large amount of data, Excel software is used to make regression analysis of the sample data and calculate the regression coefficients in this paper.

2.2 SIGNIFICANCE TEST

2.2.1 Regression equation of significance test

F criterion is used to test the linear regression equation determined initially.

Examine whether the regression coefficients of B_1 and B_2 are all zero, if they are all zero, then the linear relationship is not significant; otherwise, linear relationship is significant. The step of F criterion is as follows:

1) Suppose H_0: $B_1 = B_2 = 0$

2) Calculate the observed value of F according to the observation sample:

$$(x_{11}, x_{12}, y_1), (x_{21}, x_{22}, y_2) \ldots (x_{n1}, x_{n2}, y_n),$$

where, F is test statistic, $F = \dfrac{S_R / m}{S_e / (n - m - 1)}$, S_R is regression sum of square, S_e is residual sum of square, for binary linear regression equation $m = 2$. The calculation process in detail is omitted here.

3) Check if F meets the condition:

$F \geq F_{1-\alpha}(m, n-m-1)$, then refuse H_0, consider that linear regression effect of the linear regression equation is significance.

If F meets the condition as: $F < F_{1-\alpha}(m, n-m-1)$, then accept H_0, It means that linear regression effect of the linear regression equation is not significance.

2.2.2 Significance test of each variable factor

For the binary linear regression, it is not enough to only make test of significance of regression equation, test of significance of each regression coefficient also needs to be done. Eliminate the insignificant variable from regression equation, through test of significance of each regression coefficient and then re-establish regression equation containing all significant variable factors. Here, t criterion is used to examine whether the single variable factor is significance, respectively.

It is similar with F criterion; build the test statistic of t, and calculate the observed value of t according to the observation sample.

If t meets the condition: $|t| \geq t_{1-\alpha/2}(n-m-1)$, then consider regression variable factor x_k has significance effect on dependent variable y;

If t meets the condition: $|t| < t_{1-\alpha/2}(n-2)$, then consider regression variable factor x_k has insignificance effect on dependent variable y, then eliminate the unremarkable variable factor x_k and re-establish regression equation.

2.2.3 Correlation coefficient test

r-statistics is used to test whether the dependent variable *y* has linear correlation with variable x_k in correlation coefficient test. In the similar way like *F*-test and *t*-test. The observed value of *r* is calculated, if $|r| \geq r_\alpha(n-m-1)$, then there is linear correlation between dependent variable *y* and variable x_k; if $|r| < r_\alpha$, then there is not linear correlation between dependent variable *y* and variable x_k; and the more $|r|$ close to 1, the greater the correlation; he more $|r|$ close to 0, the smaller the correlation.

2.2.4 Determine the optimal regression equation

The method named "all out without in" is applied to determine the optimal regression equation, as the regression variables determined are only height and weight of human. That is, firstly make significance test for linear regression equation containing selected regression variables, if the equation is significance, make significance test for each regression variable, and secondly eliminate variable factor that has min effect on regression equation in non-significant factors and re-establish regression equations, thirdly, make significance test for the new equation. Do as the three steps until the optimal regression equation is obtained.

2.3 PREDICTION

It can predict the value of *y* if the fixed values of *x* are given, according to the linear regression equation established. In other words, the dependent variables of dimensions of body parts can be obtained with the indicators of height and weight according to the regression equation established with variables of height and weight. Prediction methods are point prediction and interval prediction. Point prediction can get fixed value of *y* when the x_1, x_2 are assigned according to the identified regression equation: $y = B_0 + B_1 x_1 + B_1 x_2$. Interval prediction is suitable for the situation that confidence level ($1-\alpha$) is demanded, the prediction interval of *y* with confidence level ($1-\alpha$) can be calculated according to following Equation (2):

$$(y - \sigma t_{1-\alpha/2}(n-m-1), y + \sigma t_{1-\alpha/2}(n-m-1)), \qquad (2)$$

where, *σ* is standard deviation of observation sample.

It depends on a specific situation, which method to choose (point prediction or interval prediction).

3 Determine the regression equations of Chinese adult male dimensions

Excel software is used to calculate the all parameters of the regression equation, because the size of the observed data provided by the GB10000-88 is large. The final linear regression equations of parts of human body are determined by regression significance analysis.

Eye height and shoulder height are discussed as examples to introduce the process of establishment of regression equation and prediction. The preliminary parameters related to regression equations of eye height and shoulder height are in the following Table1. Where, $\alpha = 0.05$, *n* =7, *m* =2, the results can be obtained by table lookup [11].

$r_{0.05}$ (4)=0.8114; $F_{1-0.05}$ (2,4) =6.94; $t_{1-0.05/2}$ (4) =2.7764.

For eye height: $r = 0.99989 \geq r_{0.05}(4)$ =0.8417, then eye height has significant correlation between height and weight.

As *F* =90092.52 $\geq F(2,4)$ =6.94, then the linear regression equation of eye height about height and weight is significant. As $t_1 = 42.9528 \geq t_{1-0.05/2}(4) = 2.7764$ and $t_2 = 3.5629 \geq 2.7764$, then the regression variables of height and weight are all has significant effect on eye height, and we can know that the variable of height has more significant effect than variable of weight according to the calculation result.

Finally, the regression equation of eye height with variables of height and weight is determined as follows:

$$y = 0.2077 + 0.9153x_1 + 0.5302x_2.$$

We can make prediction eye height of a Chinese adult male with his height and weight.

For example, make point prediction and interval prediction as the height and weight are assigned as 1,700 mm and 65 KG.

The eye height of point prediction is 1,590.68 mm.

When $\alpha = 0.05$, $(y - \sigma t_{1-0.05/2}(4), y + \sigma t_{1-0.05/2}(4))$ is the confidence interval of eye height, take data into the confidence interval model we can get (1587.65, 1592.35)

For shoulder height, we can know that the variable of weight is non-significant factor for regression equation of shoulder height based on the calculation shown in Table 1, for $t_2 = 1.9866 < t_{1-0.05/2}(4) = 2.7764$, and then re-establish the regression equation without variable of weight. The parameters of the correction regression equation of shoulder height are shown in the Table 1. The correction regression equation is the optimal regression equation through significance test.

The regression equations of the other parts of the body dimensions are determined as above, the regression equations of Chinese adult male dimensions are shown in Tables 2-5.

TABLE 1 Parameters for regression equations of height of eye and height of shoulder

Parameters\\Items (mm)	B_0 (Constant coefficient)	B_1 (Regression coefficient of height)	B_2 (Regression coefficient of weight)	Value of t	Value of F	r (Correlation coefficient)	σ (Standard deviation)
Eye height	0.2077	0.9153	0.5302	$t_1=42.9528$ $t_2=3.5629$	90092.52	0.999989	0.601475
Shoulder height	-68.7784	0.8352	0.5627	$t_1=20.5926$ $t_2=1.9866$	21225.36	0.999953	1.144811
Correction of shoulder height	-168.661	0.91530	0	$t_1=-163.4247$	26707.62	0.999906	1.44324

TABLE 2 Dimensions regression equations of Chinese adult male in standing posture[a]

Parameters\\Items (mm)	B_0 (Constant coefficient)	B_1 (Regression coefficient of height)	B_2 (Regression coefficient of weight)	Value of t	Value of F	r (Correlation coefficient)	σ (Standard deviation)
Eye height	0.2077	0.9153	0.5302	$t_1=42.9528$ $t_2=3.5629$	90092.52	0.999989	1.144811
Shoulder height	-168.6610	0.91530	0	$t_1=-163.4247$	26707.62	0.99906	1.44324
Elbow height	-96.1686	0.6411	0.7290	$t_1=17.4710$ $t_2=2.8450$	17175.25	0.999942	1.03578
Functional hand height	-320.1850	0.6321	0	$t_1=137.6817$	18956.25	0.999868	1.182973
Perineum high	-157.5050	0.5272	1.0485	$t_1=10.7862$ $t_2=3.0722$	7986.99	0.999875	1.37951
Tibial height	-190.2750	0.37843	0	$t_1=85.8505$	7370.31	0.999661	1.135892

[a] The linear regression equation is: $y= B_0+B_1x_1+ B_2x_2$, where, B_0, B_1, and B_2 are constant coefficient, regression coefficient of height and regression coefficient of weight; x_1 is regression variable of height, the unit of measure is millimetre; x_2 is regression variable of weight, the unit of measure is kilogram; y is the dimension of parts of human, the unit of measure is millimetre.

TABLE 3 Dimensions regression equations of Chinese adult male in sitting state[a]

Parameters\\Items (mm)	B_0 (Constant coefficient)	B_1 (Regression coefficient of height)	B_2 (Regression coefficient of weight)	Value of t	Value of F	r (Correlation coefficient)	σ (Standard deviation)
Seated height	29.4508	0.5233	0	$t_1=184.777$	34142.54	0.999927	0.729855
Cervical height, sitting	-11.2640	0.3827	0.4368	$t_1=-22.40156$ $t_2=3.661877$	28267.74	0.999965	1.44324
Eye height, sitting	-57.9047	0.51003	0	$t_1=205.8884$	42390.03	0.999941	0.638347
Shoulder height, sitting	-139.463	0.43973	0	$t_1=-198.4419$	39379.19	0.999937	0.571013
Elbow height, sitting	-348.764	0.36442	0	$t_1=183.8339$	33794.92	0.999926	0.51083
Thigh clerance height, sitting	0	42.83445	1.4376	$t_1=24.90017$	620.018	0.99599	2.130455
Knee height, sitting	-89.3428	0.3304	0.4671	$t_1=-31.89367$ $t_2=6.457367$	61190.08	0.999984	0.292406
Lower leg-foot length	36.10362	0.18772	1.0427	$t_1=11.03096$ $t_2=8.773861$	16297.06	0.999939	0.480346
buttock-popliteal length	-122.988	0.334096	0.3231	$t_1=29.40696$ $t_2=4.073074$	46652.21	0.999979	0.32068
Hip to Knee length	-59.482	0.348349	0.4835	$t_1=15.86938$ $t_2=3.154525$	15056.88	0.999934	0.619591
Lower extremity length, sitting	-257.64	0.744658	0	$t_1=107.3962$	11533.95	0.99783	1.786747

[a] the linear regression equation is: $y= B_0+B_1x_1+ B_2x_2$, where, B_0, B_1, and B_2 are constant coefficient, regression coefficient of height and regression coefficient of weight; x_1 is regression variable of height, the unit of measure is millimetre; x_2 is regression variable of weight, the unit of measure is kilogram; y is the dimension of parts of human, the unit of measure is millimetre.

TABLE 4 Horizontal dimensions regression equations of Chinese adult male[a]

Parameters Items (mm)	B_0 (Constant coefficient)	B_1 (Regression coefficient of height)	B_2 (Regression coefficient of weight)	Value of t	Value of F	r (Correlation coefficient)	σ (Standard deviation)
Chest depth	-45.6245	0.101804	1.467621	t_1=17.41333 t_2=35.94846	118350.3	0.999992	0.389786
Shoulder breadth	-238.711	0.383433	-0.51102	t_1=17.81247 t_2=-3.39955	8687.282	0.999885	0.607597
Maximum shoulder breadth	-59.9154	0.265599	0.768275	t_1=12.31861 t_2=5.102706	12616.46	0.999921	0.608576
Hip breadth	-11.0878	0.163048	0.738698	t_1=10.17592 t_2=6.601968	11697.28	0.999915	0.452264
Hip breadth, sitting	-0.5424	0.152158	1.132055	t_1=8.448848 t_2=9.001536	12652.08	0.999921	0.508334
Elbow to elbow breadth, sitting	-84.555	0.200735	2.883844	t_1=5.568218 t_2=11.45549	12045.24	0.999917	1.017552
Chest circumference	206.7272	0.215358	5.070317	t_1=12.25684 t_2=41.3238	119385.6	0.999992	0.495945
Wrist circumference	220.7002	0	8.928137	t_2=46.9323	2202.64	0.998867	7.019918
Hip circumference	159.1552	0.289085	3.943903	t_1=3.9654 t_2=7.7470	5701.406	0.999825	2.05773

[a] the linear regression equation is: $y = B_0 + B_1 x_1 + B_2 x_2$, where, B_0, B_1, B_2 are constant coefficient, regression coefficient of height and regression coefficient of weight; x_1 is regression variable of height, the unit of measure is millimetre; x_2 is regression variable of weight, the unit of measure is kilogram; y is the dimension of parts of human, the unit of measure is millimetre.

TABLE 5 Dimensions regression equations of head, hand and foot of Chinese adult male[a]

Parameters Items (mm)	B_0 (Constant coefficient)	B_1 (Regression coefficient of height)	B_2 (Regression coefficient of weight)	Value of t	Value of F	r (Correlation coefficient)	σ (Standard deviation)
Head to chin height	-46.7679	0.1534	0.208854	t_1=20.83667 t_2=4.0626	25793.95	0.999961	0.207797
Sagittal arc	-91.8649	0.2629	0	t_1=87.58131	7670.485	0.999674	0.773485
Transver-sel arc	-25.1947	0.22997	0	t_1=128.1097	16412.09	0.999848	0.462575
Head breadth	-14.6533	0.1006	0	t_1=60.39968	3648.122	0.999315	0.429352
Head length	-11.3772	0.1163	0	t_1=79.75548	6360.936	0.99607	0.375763
Head circumference	118.7744	0.2633	0	t_1=135.6043	18388.52	0.999864	0.500274
Morphological facial length	-69.9696	0.1128	0	t_1=96.07359	9230.135	0.999729	0.302615
Hand length	-48.1905	0.1377	0	t_1=105.8872	11212.1	0.999777	0.335154
Hand breadth	-30.0341	0.0668	0	t_1=65.67084	4312.659	0.999421	0.26221
Index finger length	-46.3113	0.0689	0	t_1=53.06254	2815.633	0.999113	0.334408
Foot length	-52.5146	0.1785	0	t_1=103.0215	10613.44	0.999765	0.446494
Foot breadth	-35.0796	0.0781	0	t_1=42.1762	1778.83	0.998598	0.477074

[a] the linear regression equation is: $y = B_0 + B_1 x_1 + B_2 x_2$, where, B_0, B_1, and B_2 are constant coefficient, regression coefficient of height and regression coefficient of weight; x_1 is regression variable of height, the unit of measure is millimetre; x_2 is regression variable of weight, the unit of measure is kilogram; y is the dimension of parts of human, the unit of measure is millimetre.

4 Analysis and conclusion

4.1 ACCURACY ANALYSIS OF REGRESSION EQUATION

The source of measurement data (from the Chinese human dimensions standard published in 1988) ensures the basic sample data in this study accurate and meaningful; the similarity theory of the same group human body ensures the regression equations established are reasonable. To further validate the regression equations, some Chinese male volunteers are chosen to participate in body measuring experimentation. It is found that the maximum error between actual measurement data and prediction dimension data derived from regression equations established is within 5% by comparative analysis. The theoretical analysis and actual experiment prove that the

regression equations of Chinese adult male established are accurate and have practical meaning.

4.2 CONCLUSIONS

As the measurement, data of most of the human body in static state approximately conforms to normal distribution for single-sex group. When the indicators of stature and avoirdupois are given, the measurement values of other parts of Chinese human body can be derived quickly. The result has population statistical significance for Chinese adult male with the regression equation prediction model.

The regression equations established can provide effective method to get Chinese human dimensions quickly and easily, and complements the measurement empirical formulas of Chinese human body in static state. Meanwhile, the method in this paper applies to other groups like female and minors.

5 Application

The frontal crash dummy model with Chinese civil shapes can be built in the software environment of MADYMO. When the height and weight of 50 percentile adult male are 1,690.3 millimetre and 68.8 kilogram, the body measurement data can be calculated according to equations established in this paper. Based on this, the multi rigid dummy model with Chinese body feature is built by modifying 50 percentile dummy model in MADYMO software. The Chinese simulation dummy and original dummy model in MADYMO are in Figure 1. In the figure, the original dummy model is green and as reference, Chinese simulation dummy is yellow, we can see the Chinese simulation dummy with standing posture and sitting posture.
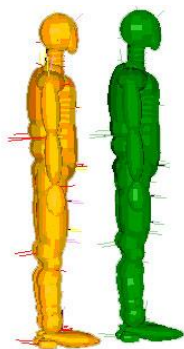
a) standing posture, the original dummy model is green, Chinese simulation dummy is yellow
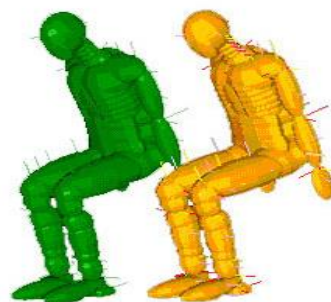
b) sitting posture, the original dummy model is green, Chinese simulation dummy is yellow

FIGURE 1 Chinese simulation dummy and original dummy model in MADYMO

References

[1] Ren J D 2010 Automotive Ergonomics *Peking University Press: Beijing (in Chinese)*
[2] GB3975-83 1983 *Anthropometric Term (in Chinese)*
[3] GB5703-85 1985 *Anthropometric Method (in Chinese)*
[4] GB10000-88 1988 *Human Dimensions of Chinese Adults (in Chinese)*
[5] Yu C Y, Lo Y H, Chiou W K 2003 The 3D scanner for measuring body surface area: a simplified calculation in the Chinese adult *Applied Ergonomics* **34** 273-78
[6] Han H, Nam Y, Choi K 2010 Comparative analysis of 3D body scan measurements and manual measurements of size Korea adult females, *International Journal of Industrial Ergonomics* **40** 530-40
[7] Wang Y J, Mok P Y, Li Y 2011 Body measurements of Chinese males in dynamic postures and application *Applied Ergonomics* **42** 900-12
[8] Bai Z H, Cao L B, Yu Z G 2008 A Research on the difference of frontal impact response between 50th percentile Chinese male and hybridiii50th percentile male *Automotive Engineering* **30**(11) 993-7 *(in Chinese)*
[9] Zhang C H, Lin D Q 2009 Parameters research on the adult anthropomorphic phantom of china *Machine* **7** 1-3 *(in Chinese)*
[10] Zhang Y R 2012 Similarity of Human-body Shape *Acta Anthropologica Sinica* **31**(3) 299-314 *(in Chinese)*
[11] Zhuang C Q 2005 Application of Mathematical Statistics *South China University of Technology Press* Guangzhou *(in Chinese)*

Authors

Xuejing Du, born in April, 1975, Tonghua City, Jilin Province, China

Current position, grades: associate professor at Northeast Forestry University.
University studies: PhD in Transportation Engineering, Jilin University, Changchun, China, 2005.
Scientific interests: computational solid mechanics, finite element analysis/optimization, fracture mechanics and numerical simulation of damage, vehicle crashworthiness optimization for crash safety design, damage progression analysis of vehicle
Publications: 2.

Huanhuan Guo, born in February, 1989, Puyang City, Henan Province, China

Current position, grades: graduate student at Northeast Forestry University,
University studies: Master degree in reading at Northeast Forestry University, Harbin, China, 2012.
Scientific interests: Vehicle crashworthiness optimization for crash safety design, damage progression analysis of vehicle, fracture mechanics and numerical simulation of damage.
Publications: 2.

Zhanyu Wang, born in October, 1975, Jian City, Jilin Province, China

Current position, grades: associate professor at Northeast Forestry University.
University studies: PhD in Vehicle Engineering, Northeast Forestry University, Harbin, China, 2006.
Scientific interests: automobile emission control & simulation, finite element analysis/optimization.
publications: 2.

Operation Research and Decision Making

Decision to rescue failed product innovation projects based on the leading innovation strategy

Xuanxuan Xu*

School of Management, Wuhan University of Technology, Wuhan, Hubei, China

Abstract

The leading innovation strategy is one of the important methods used to rescue failed product innovation projects. Rescuing failed projects based on the leading innovation strategy, rescuing time and points of competitors participating in the market can affect the probability of success. The established mathematical economic model can be used to analyse the cost of rescuing failed product innovation projects and income when competitors participate in the market in the introduction period, growth period, or mature period. The model can also be used to determine the feasibility of the leading innovation strategy and measures that enterprises should take to obtain greater profits under different competitive environments.

Keywords: leading innovation, product innovation, failed projects, rescue

1 The connotation of failed product innovation projects

According to the definition of project from the American Project Management Institute, the product innovation project is defined as a set of all types of technological activities with start and end dates to provide better products and services to customers in the prescribed time, cost, and performance parameters. It has a specific target to be completed, clear start and end dates, and limited project budget. It is an investment activity that has high risk and low success rate. Nearly 16,000 types of new products entered the market in 1991, but 90% did not achieve their business goals [1], thereby demonstrating the characteristics of product innovation projects.

Research showed that two-thirds of new products fail when they enter the market, and only less than 15% of innovative products achieve real success [2]. Numerous failed projects have brought heavy burden to enterprises. Failed projects are projects that fail to obtain the expected economic benefits and are unable to have a normal operation because of some uncertain factors. Failed product innovation projects are projects that do not reach anticipated goals. However, such projects are not worthless because they help the succeeding projects avoid risk. Knowledge gained from failed projects, as well as technical ability, will lay a solid foundation for subsequent projects [3].

2 Cost analysis of rescuing failed product innovation projects using the leading innovation strategy

Any new products with new technology have difficulty in achieving long-term endurance in the market. Enterprises must constantly research new technologies and new products to occupy or expand market shares [4]. The leading innovation strategy seeks the leading technology and products, and establishes and maintains the competitive advantages. According to Siemens, the profit of a new product will increase by 0.3% if set into production one day in advance, 1.6% if five days in advance, and 2.5% if ten days in advance. Based on failed product innovation projects, enterprises should draw from their experiences, maximize the use of the resources of failed projects, and continue product innovation using the innovation strategy to meet consumers' needs. Thus, enterprises can achieve optimal allocation of resources, differentiate the new product from other products, and determine the low cost advantage of the new product, which are important to rescue failed product innovation projects.

According to the study of Graves S. B. and Griffin A., the R&D cycle of the product D, R&D cost of the product C, and product performance Q have the following relationship: $C = C \ (Q, \ D) = Q \ (\alpha - \beta D + D^2)$, $(\alpha > 0, \beta > 0,$ and $4\alpha > \beta^2)$.

The R&D cycle D is the time from R&D to market entry for new products. The R&D cost of the product C is the sum of labour, materials, equipment, and other expenses in the process of R&D. The product performance Q is the level of the quality, function, and appearance of new products.

Whether products meet the demand of users is the key to success for projects [5]. Lei K. and Schmidt E. J. [6] believe that the uniqueness of products and degree of meeting users' demand are very important to the success or failure of product innovation projects. Enterprises must continue to research and develop new products so that the

*Corresponding author e-mail: 252434311@qq.com

Operation Research and Decision Making

performance of products can be improved after failure. The performance of a new product will reach $(Q + M)$ $(M > 0)$, and the R&D cycle will reach $(D + N)$ $(N > 0)$. In addition, some costs, including the human resource cost, machine and equipment depreciation cost, and capital cost (such as interest from the original product entering the market), should be considered in the research and development of a new product. At the same time, the product in this period can obtain a certain income. The difference between the two parts is shown in C_0. The R&D cost C is expressed as follows:　$C = C_0 + (Q + M)$ $(\alpha - \beta (D + N) + (D + N)^2)$, $(\alpha > 0, \beta > 0, \text{ and } 4\alpha > \beta^2)$.

3 Benefit analysis of rescuing failed product innovation projects using the leading innovation strategy

Enterprises need to analyse the target market and select the target customers, economy, technology, and other factors prior to researching a new product [7]. Thus, the failed project enterprises gain more experience. This experience can help enterprises promote the rate of rescuing failed product innovation projects using the leading innovation strategy. For better analysis, we assume the following:

1) Only two enterprises exist, namely, the enterprise and competitor.

2) The barriers of entry to the market are low. Enterprises can choose any strategy to rescue the failed product innovation projects.

3) No other factors influence the enterprises to use the leading innovation strategy.

3.1 ANALYSIS OF TOTAL MARKET DEMAND AND CONTRIBUTION MARGIN

3.1.1 Market demand of the new product changes with time

As shown in Figure 1, N is the R&D cycle of the new product. ρ_1 is the average growth of sales in the introduction period and $tg\,\varpi = \rho_1$. ρ_2 is the average growth of sales in the growth period and $tg\theta = \rho_2 - \rho_1$. ρ_3 is the average reduction in sales in the recession period and $tg\,\vartheta = \rho_3$. $N + r_1$ is the time the new product enters the growth period. $N + r_1 + r_2$ is the time the new product enters the mature period and $N + r_1 + r_2 + r_3$ is the time the new product enters the recession period.
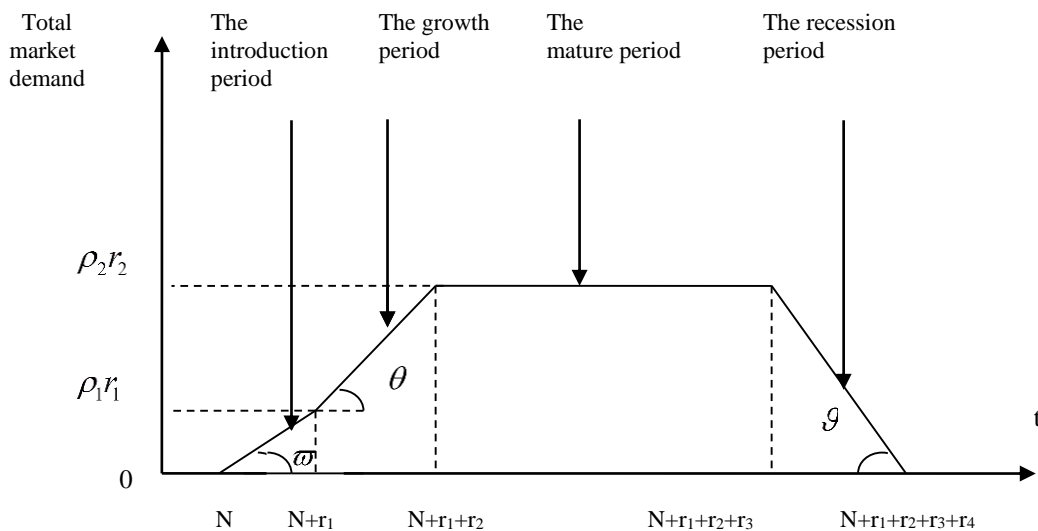

FIGURE 1 Total market demand of the new product in different periods

The total sales M of different periods are as follows:

If $N \le t \le N + r_1$, $M = \rho_1 (t - N)$;

If $N + r_1 \le t \le N + r_1 + r_2$, $M = \rho_2 (t - N - r_1)$;

If $N + r_1 + r_2 \le t \le N + r_1 + r_2 + r_3$, $M = \rho_2 r_2$;

If $N + r_1 + r_2 + r_3 \le t \le N + r_1 + r_2 + r_3 + r_4$, $M = \rho_3 (N + r_1 + r_2 + r_3 + r_4 - t)$.

3.1.2 Contribution margin of unit product in different times

Suppose that P_1 is the average price of the new product in time t, and C_1 is the average manufacturing cost of the new product in the time t. With the continuous application of the new technology, the average sales price P_1 of the new product will decline with the time t, C_1 will reduce with increasing of the production scale. According to the nature of $y = e^{-x}$, $y = e^{-x}$ will decrease with increasing production scale. According to the nature of $y = e^{-x}$, it will decrease with increasing x if $x > 0$. Thus, we suppose that P_1 and C_1 are the exponential functions of base e^{-1}.

Suppose that $P_1 = pe^{-\varepsilon t}$ ($\varepsilon > 0$) and $C_1 = ce^{-\varepsilon t}$ ($\varepsilon > 0$), where p is the unit price of the new product entering the market, c is the average manufacturing cost of the new product entering the market, and ε is the time trend coefficient.

Therefore, $R = P_1 - C_1 = (p - c) \, e^{-\varepsilon} = ve^{-\varepsilon t}$, where v is the contribution margin of the unit product entering the market.

3.2 PROFIT ANALUSIS OF RESCUING FAILED PRODUCT INNOVATION PROJECTS

3.2.1 Profit analysis when the competitor enters the market during the introduction period

Suppose that the competitor enters the market while $(N + r_5)$ $(0 \le r_5 \le r_1)$ in the introduction period because of low barriers of entry to the market. The profit of the new product can be divided into two sections in the leading innovation strategy. One is the profit T_1 before the competitor enters the market. The other is the profit T_2 after the competitor enters the market.

T_1: suppose that the enterprise has obtained the market capacity G before the competitor enters the market, and $G = \rho_1 \, (t - N)$, $(N \le t \le N + r_5)$. $R = ve^{-\varepsilon t}$. $G \times R = \rho_1 (t - N) \times ve^{-\varepsilon t} dt$. Given that G and R change with time, the profit of the enterprise before the competitor enters the market is expressed as follows:

$$T_1 = \int_N^{N+r_5} \rho_1 (t - N) \times ve^{-\varepsilon t} dt . \tag{1}$$

T_2: -after the competitor enters the market, the profit of the enterprise is: $T_2 = T_3 + T_4 + T_5 + T_6$.

According to the study of Cooper and Nakanishi, the market share of the new product is the performance function of new products of the enterprise and competitor: $S(Q, Q_c) = \gamma Q / (\gamma Q + Q_c)$. In this formula, Q is the performance of the new product of the enterprise, Q_c is the performance of the new product of the competitor, and γ is the market advantage of the enterprise to the competitor. If the performance level is the same, $Q = Q_c$.

$\gamma > 1$ indicates that the market share of the enterprise is over 50%.

If $N + r_5 \le t \le N + r_1, G = \rho_1 (t - N)$,

$$T_3 = \int_{N+r_5}^{N+r_1} \rho_1 (t - N) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt .$$

If $N + r_1 \le t \le N + r_1 + r_2, G = \rho_2 (t - N - r_1)$,

$$T_4 = \int_{N+r_1}^{N+r_1+r_2} \rho_2 (t - N - r_1) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt .$$

If $N + r_1 + r_2 \le t \le N + r_1 + r_2 + r_3, G = \rho_2 r_2$,

$$T_5 = \int_{N+r_1+r_2}^{N+r_1+r_2+r_3} \rho_2 r_2 \times S(Q, Q_c) \times ve^{-\varepsilon t} dt .$$

If $N + r_1 + r_2 + r_3 \le t \le N + r_1 + r_2 + r_3 + r_4$,

$G = \rho_3 (N + r_1 + r_2 + r_3 + r_4 - t)$,

$$T_6 = \int_{N+r_1+r_2+r_3}^{N+r_1+r_2+r_3+r_4} \rho_3 (N + r_1 + r_2 + r_3 + r_4 - t) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt.$$

Thus, the profit of the enterprise in all the periods can be expressed as follows:

$$T_1 + T_2 = T_1 + T_3 + T_4 + T_5 + T_6 = \int_N^{N+r_5} \rho_1 (t - N) \times ve^{-\varepsilon t} dt + \int_{N+r_5}^{N+r_1} \rho_1 (t - N) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt +$$

$$\int_{N+r_1}^{N+r_1+r_2} \rho_2 (t - N - r_1) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt + \int_{N+r_1+r_2}^{N+r_1+r_2+r_3} \rho_2 r_2 \times S(Q, Q_c) \times ve^{-\varepsilon t} dt + \tag{2}$$

$$\int_{N+r_1+r_2+r_3}^{N+r_1+r_2+r_3+r_4} \rho_3 (N + r_1 + r_2 + r_3 + r_4 - t) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt.$$

After R&D

$$C = C_0 + (Q + M)\left(\alpha - \beta(D + N) + (D + N)^2\right),$$

$$S(Q, Q_c) = \frac{\gamma(Q + M)}{\gamma(Q + M) + Q_c}.$$

$$\phi_1 = \int_N^{N+r_5} \rho_1 (t - N) \times ve^{-\varepsilon t} dt + \int_{N+r_5}^{N+r_1} \rho_1 (t - N) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt +$$

$$\int_{N+r_1}^{N+r_1+r_2} \rho_2 (t - N - r_1) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt + \int_{N+r_1+r_2}^{N+r_1+r_2+r_3} \rho_2 r_2 \times S(Q, Q_c) \times ve^{-\varepsilon t} dt + \tag{3}$$

$$\int_{N+r_1+r_2+r_3}^{N+r_1+r_2+r_3+r_4} \rho_3 (N + r_1 + r_2 + r_3 + r_4 - t) \times S(Q, Q_c) \times ve^{-\varepsilon t} dt - C_0 - (Q + M)\left(\alpha - \beta(N + D) + (N + D)^2\right).$$

If $\partial\phi_1 / \partial N = 0$ and $\partial\phi_1 / \partial M = 0$, the extreme value of ϕ_1 is obtained.

$$\rho_1 v \left[\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_5)}-e^{-\varepsilon N}\right)-r_5 e^{-\varepsilon(N+r_5)} \right] + \rho_1 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[r_1 e^{-\varepsilon(N+r_1)}-r_5 e^{-\varepsilon(N+r_5)}+\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1)}-e^{-\varepsilon(N+r_5)}\right) \right]+$$

$$\rho_2 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[r_2 e^{-\varepsilon(N+r_1+r_2)}+\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1+r_2)}-e^{-\varepsilon(N+r_1)}\right) \right]+$$

$$\rho_2 r_2 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[e^{-\varepsilon(N+r_1+r_2+r_3)}-e^{-\varepsilon(N+r_1+r_2)} \right]-(Q+M)(2N+2D-\beta)+ \tag{4}$$

$$\rho_3 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1+r_2+r_3)}-e^{-\varepsilon(N+r_1+r_2+r_3+r_4)}\right)-r_4 e^{-\varepsilon(N+r_1+r_2+r_3)} \right]=0.$$

and

$$\rho_1 v \left[\frac{r_5}{\varepsilon}e^{-\varepsilon(N+r_5)}-\frac{r_1}{\varepsilon}e^{-\varepsilon(N+r_1)}+\frac{1}{\varepsilon^2}\left(e^{-\varepsilon(N+r_5)}-e^{-\varepsilon(N+r_1)}\right) \right]\left(\frac{Q_c r}{\left[r(Q+M)+Q_C \right]^2} \right)+$$

$$+\rho_2 v \left[-\frac{r_2}{\varepsilon}e^{-\varepsilon(N+r_1+r_2)}+\frac{1}{\varepsilon^2}\left(e^{-\varepsilon(N+r_1)}-e^{-\varepsilon(N+r_1+r_2)}\right) \right]\left(\frac{Q_c r}{\left[r(Q+M)+Q_C \right]^2} \right)+$$

$$+\frac{\rho_2 r_2 v}{-\varepsilon}\left[e^{-\varepsilon(N+r_1+r_2+r_3)}-e^{-\varepsilon(N+r_1+r_2)} \right]\left(\frac{Q_c r}{\left[r(Q+M)+Q_C \right]^2} \right)-\left(\alpha-\beta(N+D)+(N+D)^2 \right)+ \tag{5}$$

$$+\rho_3 v \left[\frac{r_4}{\varepsilon}e^{-\varepsilon(N+r_1+r_2+r_3)}+\frac{1}{\varepsilon^2}\left(e^{-\varepsilon(N+r_1+r_2+r_3+r_4)}-e^{-\varepsilon(N+r_1+r_2+r_3)}\right) \right]\left(\frac{Q_c r}{\left[r(Q+M)+Q_C \right]^2} \right)=0.$$

When the R&D cycle and performance meet the above two formulas, the enterprise can obtain maximum profits.

3.2.2 Profit analysis when the competitor enters the market during the growth period

When the competitor enters the market in the growth period, the profit is as follows:

$$\phi_2 = \int_N^{N+r_1} \rho_1 r_1 \times v e^{-\varepsilon t}dt + \int_{N+r_1}^{N+r_1+r_5} \rho_2 \left(t-N-r_1\right)\times v e^{-\varepsilon t}dt +$$

$$\int_{N+r_1+r_5}^{N+r_1+r_2} \rho_2 \left(r_2-r_5\right)\times S(Q,Q_c)\times v e^{-\varepsilon t}dt + \int_{N+r_1+r_2}^{N+r_1+r_2+r_3} \rho_2 r_2 \times S(Q,Q_c)\times v e^{-\varepsilon t}dt + \tag{6}$$

$$\int_{N+r_1+r_2+r_3}^{N+r_1+r_2+r_3+r_4} \rho_3 \left(N+r_1+r_2+r_3+r_4-t\right)\times S(Q,Q_c)\times v e^{-\varepsilon t}dt - C_0 - (Q+M)\left(\alpha-\beta(D+N)+(D+N)^2 \right).$$

If $\partial\phi_2/\partial N = 0$ and $\partial\phi_2/\partial M = 0$, the extreme value of ϕ_2 is obtained.

$$\rho_1 v \left[r_1 e^{-\varepsilon(N+r_1)}+\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1)}-e^{-\varepsilon N}\right) \right] + \rho_2 v \left[r_2 e^{-\varepsilon(N+r_1+r_5)}+\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1+r_5)}-e^{-\varepsilon(N+r_1)}\right) \right]+$$

$$\rho_2 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[r_2\left(e^{-\varepsilon(N+r_1+r_2+r_3)}-e^{-\varepsilon(N+r_1+r_5)}\right)-r_5\left(e^{-\varepsilon(N+r_1+r_2)}-e^{-\varepsilon(N+r_1+r_5)}\right) \right]- \tag{7}$$

$$-\rho_3 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c}\left[\frac{1}{\varepsilon}\left(e^{-\varepsilon(N+r_1+r_2+r_3+r_4)}-e^{-\varepsilon(N+r_1+r_2+r_3)}\right)-r_4 e^{-\varepsilon(N+r_1+r_2+r_3)} \right]-(Q+M)(2N+2D-\beta)=0.$$

and

$$\rho_2 v \left[\frac{r_2 - r_5}{\varepsilon} \left(e^{-\varepsilon(N+r_1+r_5)} - e^{-\varepsilon(N+r_1+r_2)} \right) + \frac{r_2}{\varepsilon} \left(e^{-\varepsilon(N+r_1+r_5)} - e^{-\varepsilon(N+r_1+r_2+r_3)} \right) \right] \left(\frac{Q_c r}{\left[r(Q+M)+Q_c \right]^2} \right) +$$

$$+ \rho_3 v \left[\frac{r_4}{\varepsilon} e^{-\varepsilon(N+r_1+r_2+r_3)} + \frac{1}{\varepsilon^2} \left(e^{-\varepsilon(N+r_1+r_2+r_3+r_4)} - e^{-\varepsilon(N+r_1+r_2+r_3)} \right) \right] \left(\frac{Q_c r}{\left[r(Q+M)+Q_c \right]^2} \right) - \left(\alpha - \beta(N+D) + (N+D)^2 \right) = 0. \quad (8)$$

When the R&D cycle and performance meet the above two formulas, the enterprise can obtain maximum profits.

3.2.3 Profit analysis when the competitor enters the market during the mature period

When the competitor enters the market in the mature period, the profit is as follows:

$$\phi_3 = \int_N^{N+r_1} \rho_1 (t-N) \times ve^{-\varepsilon t} dt + \int_{N+r_1}^{N+r_1+r_2} \rho_2 (t-N-r_1) \times ve^{-\varepsilon t} dt + \int_{N+r_1+r_2}^{N+r_1+r_{2+}r_5} \rho_2 r_2 \times ve^{-\varepsilon t} dt +$$

$$\int_{N+r_1+r_2+r_5}^{N+r_1+r_2+r_3} \rho_2 r_2 \times S(Q,Q_c) \times ve^{-\varepsilon t} dt + \int_{N+r_1+r_2+r_3}^{N+r_1+r_2+r_3+r_4} \rho_3 (N+r_1+r_2+r_3+r_4-t) \times S(Q,Q_c) \times ve^{-\varepsilon t} dt - \quad (9)$$

$$C_0 - (Q+M) \left(\alpha - \beta(N+D) + (N+D)^2 \right).$$

If $\partial\phi_3 / \partial N = 0$ and $\partial\phi_3 / \partial M = 0$, the extreme value of ϕ_3 is obtained.

$$\rho_1 v \left[r_1 e^{-\varepsilon(N+r_1)} + \frac{1}{\varepsilon} \left(e^{-\varepsilon(N+r_1)} - e^{-\varepsilon N} \right) \right] + \rho_2 v \left[r_2 e^{-\varepsilon(N+r_1+r_2)} + \frac{1}{\varepsilon} \left(e^{-\varepsilon(N+r_1+r_2)} - e^{-\varepsilon(N+r_1)} \right) \right] +$$

$$\rho_2 r_2 v \left[e^{-\varepsilon(N+r_1+r_2+r_5)} - e^{-\varepsilon(N+r_1+r_2)} \right] + \rho_2 r_2 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c} \left[e^{-\varepsilon(N+r_1+r_2+r_3)} - e^{-\varepsilon(N+r_1+r_2+r_5)} \right] -$$

$$\rho_3 v \frac{\gamma(Q+M)}{\gamma(Q+M)+Q_c} \left[\frac{1}{\varepsilon} \left(e^{-\varepsilon(N+r_1+r_2+r_3+r_4)} - e^{-\varepsilon(N+r_1+r_2+r_3)} \right) - r_4 e^{-\varepsilon(N+r_1+r_2+r_3)} \right] + \quad (10)$$

$$(Q+M)(2N+2D-\beta) = 0.$$

and

$$- \frac{\rho_2 r_2 v}{\varepsilon} \left[e^{-\varepsilon(N+r_1+r_2+r_3)} - e^{-\varepsilon(N+r_1+r_2+r_5)} \right] \left(\frac{Q_c r}{\left[r(Q+M)+Q_c \right]^2} \right) - \left(\alpha - \beta(N+D) + (N+D)^2 \right)$$

$$+ \rho_3 v \left[\frac{r_4}{\varepsilon} e^{-\varepsilon(N+r_1+r_2+r_3)} + \frac{1}{\varepsilon^2} \left(e^{-\varepsilon(N+r_1+r_2+r_3+r_4)} - e^{-\varepsilon(N+r_1+r_2+r_3)} \right) \right] \left(\frac{Q_c r}{\left[r(Q+M)+Q_c \right]^2} \right) = 0. \quad (11)$$

When the R&D cycle and performance meet the above two formulas, the enterprise can obtain maximum profits.

4 Conclusions

By analysing the profit in these three cases, enterprises can examine different situations while rescuing. Enterprises can analyse the feasibility of the leading innovation strategy, as well as different measures to confront competitors using the leading innovation strategy.

1) $\phi_1 > 0$, $\phi_2 > 0$, or $\phi_3 > 0$ shows that rescuing the failed product innovation project in the leading innovation strategy is profitable. For the maximized value, the

enterprise will analyse the profits of different investment activities to determine the optimal investment scheme, and decide whether to rescue the failed project. Moreover, introducing the concept of rescuing costs in the model, which is related to s rescuing time, may help enterprises determine the best rescuing chance according to the final conclusion of the model.

2) $\phi_1 < \phi_2 < \phi_3$ or $\phi_1 < \phi_3 < \phi_2$ shows that the greatest effect to the profit is observed when the competitor enters the market during the introduction period. In this case, enterprises can increase the intensity of research and development of innovative products, promote the degree of differentiation of the product, reduce costs, maintain a high market threshold to entry, and decrease the speed of

Operation Research and Decision Making

other enterprises entering the market to obtain more profits.

3) $\phi_2 < \phi_1 < \phi_3$ or $\phi_2 < \phi_3 < \phi_1$ shows that the greatest effect to the profit is observed when the competitor enters the market during the growth period. In the growth period, enterprises can quickly grab the market share by innovation products. If a competitor enters the market in this period, it will change the pattern of the market and influence the goal of expanded market share within a short time period. Only by accumulating adequate resources and shortening the growth period of innovative products can the enterprise reduce the negative effects from a competitor.

4) $\phi_3 < \phi_1 < \phi_2$ or $\phi_3 < \phi_2 < \phi_1$ shows that the greatest effect to the profit is observed when the competitor enters the market during the mature period. An enterprise in the mature period has formed a relatively stable customer base. If the competitor chooses to join the market at this time, the risk and probability of failure are high. Therefore, the probability of $\phi_3 < \phi_1 < \phi_2$ or $\phi_3 < \phi_2 < \phi_1$ is low.

In the different life cycles, the internal and external environments of enterprises differ, so an optimal innovation model does not exist. Enterprises need to create new products constantly to meet the needs of customers because of unstable customer demand. Even though a competitor enters the market, enterprises should continue to innovate, seek for the best balance point in the R&D cycle, and develop product performance, and research development costs and benefits to obtain the largest profit. In view of failed product innovation projects, rescuing time and rescuing strategy are the keys to success.

References

[1] Balachandra R, Friar J H 1997 *IEEE Transaction on Engineering Management* **44**(3) 276-87

[2] Chen J,Shang P 2003 The strategic management of product innovation project *Science & technology progress and policy* **20**(4) 23-5

[3] Gao J 2011 The real options incentive to"failed" R&D projects *Science Research Management* **32**(8) 127-32 *(in Chinese)*

[4] Luo Y 2007 Study on Product Innovation Based on Deeply Development of Implicit Demand – View of Information Cognition *Science of science and management of S&T* **28**(5) 82-7 *(in Chinese)*

[5] Zhao J, Chen J 1991 Statistics analysis on the successful or failing factors of technological in notation project *Chinese journal of industrial engineering/engineering management* **5**(2) 57-60 *(in Chinese)*

[6] Cooper R G, Kleinschmidt E J 1993 New Product Success In The Chemical Industry *Industrial Marketing Management* **22**(2) 85-99

[7] Ji G 2012 Optimal decision on innovation effort and pricing under the constraint of maximum willingness to pay *Science of science and management of S&T* **33**(11) 38-45 *(in Chinese)*

Authors

Xuanxuan Xu, born in August, 1985, Hubei Province, China

Current position, grades: PhD.
University studies: Master of Management at Southwest University for Nationalities in China.
Scientific interests: project management, risk management.
Publications: more than 9 papers.

Mechanism of formation in the selling price of urban industrial land in China based on the multi-attribute auction theory

Jian-gang Shi, Jin-can Liu*

Yuntong Building 509, School of Economics & Management, Tongji University, 1239 Siping Road, Yangpu District, Shanghai China, 200092

Abstract

There is the fierce competition for investment introduction among local governments in China. Thus, the auction relationship about urban industrial land belongs to a kind of reverse auction between local government and investment enterprise. To explore the mechanism of formation in the selling price of urban industrial land, we build a multi-attribute first-score sealed-bid auction model. Our studies reveal that the selling price of urban industrial land is the decreasing function of local government's expected comprehensive earnings from the investment introduction, the intensity of competition for investment introduction among local governments or its cost factors of providing some attributes. However, it is the increasing function of enterprise's weights for different attributes. They appropriately explain why the selling price of urban industrial land has been at the low level for a long time in China. Yet, the comprehensive earnings and the realization of land asset value are not a pair of irreconcilable conflicts. Local governments should prioritize the attributes, which the investment enterprises care about, then reduce the cost of providing these attributes. In the end, it will resolve the interest conflicts in the process of industrial land operation.

Keywords: industrial park, industrial land, selling price, multi-attribute auction

1 Introduction

The urban land is owned by the local governments in China. The ownership of the urban land is forbidden to merchandise. However, it is the land use rights for a certain period that are traded in the land market. According to the Chinese law, the maximum selling period of the urban industrial land is 50 years. Local governments have to sell the use rights of their land based on this law. In this sense, the selling price of the land is equivalent to the selling price of the land use rights. Based on the average selling price levels of the urban lands mainly monitored in China from 2001 to 2011, the selling price of commercial land rose by 2.4 times; residential land by 3.68 times. However, the selling price of urban industrial lands only rose by 0.44 times. When compared with the absolute values of the average selling price levels in 2011, the urban industrial land was only 11.5% of the commercial land and 14.4% of the residential land. Even though the selling price of industrial land in Shanghai was far higher than the national average level, it was only 45% of that in Shanghai in 2011. From 2001 to 2011, the industrial land price level rose by 0.63 times in Shanghai, which was also far higher than the national average.

Therefore, the changes in selling price of urban industrial land are significantly different from those of the commercial land and the residential land. The formation of selling price of the urban industrial land in China has its unique mechanism, which is possibly associated with the competition for investment introduction among local governments and the land-related properties besides the price.

To understand the mechanism of formation in the selling price of urban industrial land in China, Wu analysed the relationship between the low selling price and the earnings from the investment introduction by building a game model [1]. Zhang studied the mechanism from the view of the competition among local governments in China [2]. Xu explored the forming process and mechanism of industrial land selling price based on the game model and cost-benefit analysis method [3]. Yang analysed why the low selling price of industrial land was caused by the failure of local governments' roles in China [4]. Wang explored the factors affecting the selling prices of urban industrial land, such as industrial agglomeration and government intervention based on a panel data of 35 Chinese cities from 2000 to 2010 [5]. Liu investigated the city-level impact of the economic fundamentals on the selling prices in China by using the panel data model [6].

Now we focus on the mechanism of formation in selling price of urban industrial land by using the multi-attribute auction theory in the context of fierce competition for investment introduction among local governments in China.

2 The multi-attribute auction theory

Multi-attribute auction is an auction mode that the auctioneer and bidder have multiple negotiations on price and other attributes [7]. For example, in a procurement process, the buyer (auctioneer) announces the desired characteristics of a particular item, which may include the

Corresponding author e-mail: liujincan@sina.com

Operation Research and Decision Making

price, quality, delivery deadline, and warranty period. A fixed number of n sellers (bidder) who decide to send bids have to specify the full configuration. Finally, the buyer decides which bid it prefers, and the agent who suggests this bid, is called the winner agent and committed to providing it. Multi-attribute auction is an evolution of traditional auction, mainly suitable for the reverse auction centred on the buyer. It fully satisfies a variety of individual demands of the auctioneer, and is more in line with the real economic activity.

Che put forward a two-dimensional multi-attribute auction model, and studied several deformation auction protocols under sealed auction mechanism [8]. On the basis of Che's research, Brallco studied further the influence of the related costs for the bidders on the auction results [9]. Parkes and Kalagnanam analysed the multi-attribute auction of multi-rounds with the linear programming method [10]. David studied systematically multi-attribute auction [11-13]. In addition to the above theoretical researches, Bichler and Chen-Ritzo compared single-attribute auction with multi-attribute auction by the method of experimental economics, and found out that multi-attribute auction can improve both the buyer's and the seller's utilities. Chinese scholars gave priority to theoretical research, mainly focusing on the optimal mechanism design of multi-attribute auction, such as the optimal bidding strategy for bidders and the system design

3 The characteristics of the urban industrial land auction in China

As a developing country, there is the fierce competition of investment introduction among local governments in China. This kind of competition is beneficial to the investors when they are in the game with local governments. As is well known, when an enterprise chooses its investment place, it considers not only the industrial land selling price, but also such factors as location, industry development, tax level and social management. If having several invitations from local governments, an enterprise will have the right to choose investment place. It can select the most suitable region to set up a company according to its own revenue function. Under such circumstances, the negotiations about investment place between investment enterprise and multiple local governments are a multi-attribute first-score sealed-bid auction. Investment enterprise, as auctioneer, has its own scoring function; while local governments, as bidders, bid on the land-related attributes including the selling price. In the end, the investment enterprise decides which place it prefers, and the local government that suggested this bid, is committed to offer it and wins the enterprise to invest in the local area, as illustrated in Figure 1.
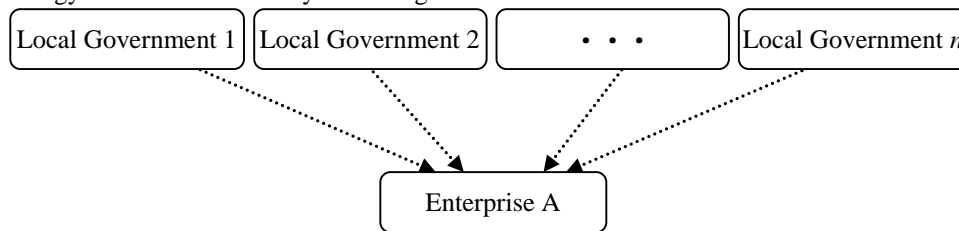
about online multi-attribute auction, etc.

FIGURE 1 The auction relationship between enterprise and local governments

4 The multi-attribute auction model about the urban industrial land in China

In the multi-attribute auction model about the urban industrial land in China, the buyer of land (investment enterprise) is the sponsor of the auction, the seller of land (local government) is the bidder. Local governments carry on bidding competition on the multiple attributes that investment enterprise requires. In the study of multi-attribute auction, though each model is not identical, they all consist of five functions associated with the buyer or the seller:

1) the bidder's cost function;
2) the bidder's utility function;
3) the auctioneer's utility function;
4) the auctioneer's scoring function;
5) the auctioneer's announced bid's value function.

Based on the characteristics of industrial land multi-attribute auction, the modelling assumption and the corresponding functions are provided in the following sub-sections.

4.1 MODELING ASSUMPTIONS

Both local government and enterprise are risk neutral and the multiple attributes related to the industrial land are utility independent. Given this assumption, we can use the additive weighting utility function to combine the different attributes into a decision rule.

Both local government and enterprise are rational, and the game between local governments is non-cooperative. This hypothesis suggests that local government and enterprise will try to maximize their utility and will not take any action that brings them negative utility. Moreover, the local governments are rivals with each other, so there is no cooperation or conspiracy among them.

Local government's cost parameter θ_i is an independently and identically distributed function $F(\theta_i)$ over $\left[\underline{\theta}, \overline{\theta}\right]$ $(0 \leq \underline{\theta_i} \leq \theta_i \leq \infty)$, $F(\underline{\theta_i}) = 0$, $F(\overline{\theta_i}) = 1$, for which a positive, continuously differentiable density

function $f(\theta_i)$ exists. This is the common knowledge for local governments and investment enterprise. Such assumption is similar to the retention value assumptions in the traditional single price auction mechanism. Because of complete symmetry among local governments, the subscript i will be omitted in the rest of the paper.

The industrial land is of independent private values for local governments. The local government can clearly know its earnings from investment introduction according to the exact value of cost parameters, but cannot know how many earnings other local governments will get.

Transaction cost is considered as zero here. In the competition of investment introduction, information cost, participation cost and opportunity cost do exist. In order to simplify the complexity of model, this paper does not however take into account of the transaction cost.

4.2 THE DESCRIPTION OF THE PROTOCOL AND THE UTILITY FUNCTIONS

Multi-attribute auction focuses mainly on the relationship between price and other attributes. When analysing a case of multi-attribute auction, there are an arbitrary number of attributes $(m + 1)$. We can divide all these attributes into two categories: One is the price P, the other are the attributes besides price. Given the assumptions mentioned above, if the interest features of local government and enterprise are taken into account, the two sides' utility functions can be expressed as follows:

1) Local government's utility function (U_s):

$$U_s = P + R_s - \theta \sum_{i=1}^{m} \alpha_i q_i^{k_i} . \tag{1}$$

R_s represents the comprehensive earnings of local government from investment introduction, including tax, employment, etc. There are fixed coefficients for each of the attributes, namely, α_i is the cost coefficient of the attribute q_i. k_i is the index of the attribute q_i, and $k_i > 1$, this means that as q_i increases, local government's cost and marginal cost of providing the attribute q_i both increase.

2) Local government's cost function (C_s):

In the multi-attribute auction theory, cost function is opposite to price, so the Equation (1) can be reformed as the Equation (2):

$$U_s = P - \left(\theta \sum_{i=1}^{m} \alpha_i q_i^{k_i} - R_s \right). \tag{2}$$

Thus, local government's cost function can be expressed as follows:

$$C_s = \theta \sum_{i=1}^{m} \alpha_i q_i^{k_i} - R_s . \tag{3}$$

This equation indicates that for local government, the comprehensive earnings it can obtain mean a reduction to the cost of providing all attributes.

3) Enterprise's utility function (U_b):

$$U_b = \sum_{i=1}^{m} \beta_i q_i - P . \tag{4}$$

Enterprise's utility is the quasi-linear function of the attribute q_i, β_i is the weights associated with q_i, which indicates enterprise's preference for different attributes.

Given enterprise's utility function, it will announce a scoring function, which is used for choosing among bids. If the number of bidders is large enough, the auctioneer can announce its true preferences [13]. Under the circumstance of the fierce investment introduction, the scoring function of enterprise is considered as the same with its real utility function in the multi-attribute first-score sealed-bid auction. That is:

$$S_b = \sum_{i=1}^{m} \beta_i q_i - P . \tag{5}$$

From the scoring function, we can infer that the announced value of industrial land for the enterprise is:

$$V_b = \sum_{i=1}^{m} \beta_i q_i . \tag{6}$$

4.3 THE OPTIMAL ATTRIBUTE VALUES OF INDUSTRIAL LAND FOR LOCAL GOVERNMENT

In the case of multi-attribute auction, the local government faces the challenge of how to bid to maximize its expected utility. If the seller is rational, in order to determine the optimal non-price attribute values maximizing the seller's utility, we need to introduce the following important lemma:

Lemma 1: In a multi-attribute auction protocol, given the scoring rule and the seller's utility function, the optimal non-price attributes q_i that maximize the seller's utility can be chosen at $q_i^*(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, where,

$$q_i^*(\theta) = \arg\max_{q_i} \{V_b(q_1, \cdots, q_m) - C_s(q_1, \cdots, q_m, \theta)\}. \tag{7}$$

The optimal attribute values are calculated independently of the price and the seller's beliefs about the other participants. This proof can be found in David's thesis [12]. Lemma 1 is critical for solving the bidder's optimal bidding strategy of multi-attribute auction. Because the choices of the optimal non-price attributes are independent of the price, the multi-attribute auction will be turned into traditionally single price attribute auction by solving the optimal non-price attribute values first.

According to the lemma 1, the differences between

enterprise's announced bid's value V_b and local government's cost function C_s can be shown in Equation (8).

$$y = V_b - C_s = R_s + \sum_{i=1}^{m} \beta_i q_i - \theta \sum_{i=1}^{m} \alpha_i q_i^{k_i} \qquad (8)$$

The optional attribute values $q_i^*(\theta)$ of industrial land which maximize y, should satisfy Equation (9):

$$\frac{\partial y}{\partial q_i}\Big|_{q_i^*} = \beta_i - \theta \alpha_i s_i q_i^{(k_i-1)} = 0 \qquad (9)$$

Through solving Equation (9), the optimal attribute values of industrial land are shown as Equation (10):

$$q_i^*(\theta) = \left[\frac{\beta_i}{\theta \alpha_i k_i}\right]^{\frac{1}{k_i-1}}, \qquad (10)$$

where, $i = 1,2,...,m$; $k_i > 1$.

4.4 THE OPTIMAL SELLING PRICE OF INDUSTRIAL LAND FOR LOCAL GOVERNMENT

After calculating each optimal attribute value, what local government needs to do is to determine the selling price of industrial land. Thus, the multi-attribute auction of industrial land is turned into a typical single price attribute auction. Under the first-score sealed-bid mechanism, the optional selling price of industrial land for a local government can be calculated by using Che's method [8], as follows:

$$P^*(\theta) = C_s\left(q_1^*(\theta),...,q_n^*(\theta),\theta\right) +$$

$$\int_{\theta}^{\bar{\theta}} C_{s\theta}\left(q_1^*(t),...,q_n^*(t),\theta\right)\left[\frac{1-F(t)}{1-F(\theta)}\right]^{n-1} dt,$$

where, $C_{s\theta}(\cdot) = \frac{\partial C_s}{\partial \theta}$. Here, local government's cost parameter θ is considered to subject to independent, identical and uniform distribution over $\left[\underline{\theta},\bar{\theta}\right]$. Substitute

$F(\theta) = \frac{\theta-\underline{\theta}}{\bar{\theta}-\underline{\theta}}$ and $q_i^*(\theta) = \left[\frac{\beta_i}{\theta \alpha_i k_i}\right]^{\frac{1}{k_i-1}}$ into the equation

above, then the optional selling price of industrial land for local government is:

$$P^*(\theta) = \theta \sum_{i=1}^{m} \alpha_i \left[\frac{\beta_i}{\theta \alpha_i k_i}\right]^{\frac{k_i}{k_i-1}} - R_s +$$

$$\sum_{i=1}^{m} \alpha_i \left[\frac{\beta_i}{\alpha_i k_i}\right]^{\frac{k_i}{k_i-1}} \frac{1}{\left(\bar{\theta}-\theta\right)^{n-1}} \int_{\theta}^{\bar{\theta}} \frac{\left(\bar{\theta}-t\right)^{n-1}}{(t)^{\frac{k_i}{k_i-1}}} dt = \qquad (11)$$

$$\sum_{i=1}^{m} \alpha_i \left[\frac{\beta_i}{\alpha_i k_i}\right]^{\frac{k_i}{k_i-1}} \left[\frac{1}{\theta^{\frac{1}{k_i-1}}} + \frac{1}{\left(\bar{\theta}-\theta\right)^{n-1}} \int_{\theta}^{\bar{\theta}} \frac{\left(\bar{\theta}-t\right)^{n-1}}{(t)^{\frac{k_i}{k_i-1}}} dt\right] - R_s.$$

From the Equation (11), we know that the optional selling price of industrial land for a given local government is affected by its cost parameters, the estimation of cost parameters for other local governments, as well as the scoring function announced by the enterprise. Thus, there are the following properties regarding the optional selling price for a given local government.

1) The optional selling price of industrial land for a given local government is the decreasing function of its expected comprehensive earnings from the investment introduction. The more the expected comprehensive earnings are, the lower demand for the selling price will be.

2) Where

$$\frac{1}{\left(\bar{\theta}-\theta\right)^{n-1}} \int_{\theta}^{\bar{\theta}} \frac{\left(\bar{\theta}-t\right)^{n-1}}{(t)^{\frac{k_i}{k_i-1}}} dt = \int_{\theta}^{\bar{\theta}} \left(\frac{\bar{\theta}-t}{\bar{\theta}-\theta}\right)^{n-1} \frac{1}{(t)^{\frac{k_i}{k_i-1}}} dt \qquad \text{and}$$

$\frac{\bar{\theta}-t}{\bar{\theta}-\theta} < 1$, so $P^*(\theta)$ is the decreasing function of n. This indicates that the competitive intensity for investment introduction among local governments significantly influences the selling price of industrial land. The fiercer the competition among local governments is, the lower the demand for the selling price will be.

3) $P^*(\theta)$ is the increasing function of β_i. This means: if some land-related attributes q_i is valued by investment enterprise, local governments will provide more such attributes q_i to cater to enterprise's demand. Obviously, this will lead to cost addition for a local government. The local government has to improve the selling price of industrial land to get some earnings.

4) To transform the Equation (11) into the Equation (12):

$$P^*(\theta) =$$

$$\sum_{i=1}^{m} \frac{1}{\alpha_i^{\frac{1}{k_i-1}}} \left(\frac{\beta_i}{k_i}\right)^{\frac{k_i}{k_i-1}} \left[\frac{1}{\theta^{\frac{1}{k_i-1}}} + \frac{1}{\left(\bar{\theta}-\theta\right)^{n-1}} \int_{\theta}^{\bar{\theta}} \frac{\left(\bar{\theta}-t\right)^{n-1}}{(t)^{\frac{k_i}{k_i-1}}} dt\right] - R_s. \qquad (12)$$

Given $k_i > 1$, then $\frac{1}{k_i-1} > 0$, so we know that $P^*(\theta)$ is

the decreasing function of α_i. According to $\frac{\partial P^*(\theta)}{\partial \theta} < 0$,

it can be inferred that $P^*(\theta)$ is also the decreasing function of θ, which was already proven in Zheng's thesis [14].

Cost factors (α_i and θ) for a local government both negatively impact on the selling price of industrial land. When its cost factors of providing some land-related attributes are bigger, it will provide fewer amounts of such attributes. The local government has to lower the selling price of industrial land to increase its desired probability of investment introduction.

5 Example

To illustrate the real situation of urban industrial parks in China, we now assume that a foreign enterprise (b) is trying to select a city to set up a factory. There are three local governments (s_1, s_2, s_3) competing for the settlement of such a factory in their respective cities. The foreign enterprise may conduct a reverse auction against the three local governments. Consequently, the foreign enterprise becomes the auctioneer, and the three local governments are the bidders. Here, we assume that the attributes which the enterprise considers are the industrial agglomeration level q_1, the infrastructure condition q_2, and the management environment q_3 in their industrial parks. At the same time, we assume the utility functions of the participants are as follow:

$$U_b(p, q_1, q_2, q_3) = -P + 2 \cdot q_1 + 4 \cdot q_2 + 6 \cdot q_3,$$

$$U_{s_1}(q_1, q_2, q_3) = P_{s_1} + R_{s_1} - 0.2 \cdot (1 \cdot q_1^2 + 2 \cdot q_2^2 + 3 \cdot q_3^2),$$

$$U_{s_2}(q_1, q_2, q_3) = P_{s_2} + R_{s_2} - 0.2 \cdot (1 \cdot q_1^2 + 3 \cdot q_2^2 + 5 \cdot q_3^2),$$

$$U_{s_3}(q_1, q_2, q_3) = P_{s_3} + R_{s_3} - 0.2 \cdot (1 \cdot q_1^2 + 4 \cdot q_2^2 + 7 \cdot q_3^2).$$

Since the three local governments decide about their bids based on the enterprise's scoring function, here, let the scoring function still be consistent with the enterprise's utility function:

$$S_b(p, q_1, q_2, q_3) = -P + 2 \cdot q_1 + 4 \cdot q_2 + 6 \cdot q_3. \quad (13)$$

Given the information above, under the first-score sealed-bid auction protocol, we can calculate the optimal attribute values of industrial land, the selling prices of land, and the scoring values of three local governments respectively based on Equations (10), (11) and (13). List them separately as follows:

$$BID_1 = (P = 44.8 - R_{s_1}, q_1 = 5, q_2 = 5, q_3 = 5),$$

$$BID_2 = (P = 30.8 - R_{s_2}, q_1 = 5, q_2 = 3.3, q_3 = 3),$$

$$BID_3 = (P = 24.51 - R_{s_3}, q_1 = 5, q_2 = 2.5, q_3 = 2.14),$$

$$S_1 = 15.2 + R_{s_1}; S_2 = 10.4 + R_{s_2}; S_3 = 8.33 + R_{s_3}.$$

From the example above, we can learn that as the cost coefficient decreases, the local government may provide this type of attribute more. Thus it will require the higher selling price of industrial land. According to the calculated results, we can also learn that the selling price of industrial land is directly affected by the comprehensive earnings brought by the introduction of an enterprise. The higher the comprehensive earnings are, the lower the demand for the selling price is. Then the final bidding result will be affected accordingly. When $R_{s_1} = R_{s_2} = R_{s_3}$, that is, the three local governments can get the same comprehensive earnings from the investment introduction, the winner in this case will be the local government s_1 who obtains a score equal to S_1.

6 Conclusion

We explore the mechanism of formation in the selling price about urban industrial land in China based on the theory of multi-attribute first-score sealed-bid auction. Our results show that the competition for the investment introduction among local governments will significantly influence the selling process of industrial land. When faced with an alternative between the realization of land asset value and the comprehensive earnings from the investment introduction, local governments have to make efforts to win the competition for the investment introduction at the expense of land asset value in order to maximize the overall revenues. If there are more and more local governments to participate in the auction, and the enterprise possesses the strong negotiation ability, local governments will be forced to provide a number of preferential policies including land price. This will make each other fall into the predicament of "competition hitting new lows". Obviously, such results are in line with the basic characteristics of the transaction of urban industrial land in China. They explain why the selling price of urban industrial land has been at the low level for a long time.

Our results also indicate that the comprehensive earnings and the realization of land asset value are not a pair of irreconcilable conflicts. Local governments need to differentiate the attributes, which the investment enterprises more care about first, such as industrial agglomeration level, infrastructure construction level, management environment etc. They then reduce the cost of providing these attributes through scientific planning and strict management if possible, so as to improve the amount of these attributes and to obtain the higher selling price of industrial land. In this way, local government cannot only improve the successful rate of investment introduction, it also can better the market value of industrial land gradually. Thereby, local government can resolve the interest conflicts in the process of industrial land operation, and make land operation as well as industry operation develop in coordinate with each other.

References

[1] Wu Y 2007 Regional equilibrium of industrial land price and management strategy based on game theory *Journal of Zhejiang University (humanities and social sciences)* **37**(4) 124-33 *(in Chinese)*

[2] Zhang Q 2006 Local governments' competition and the leasing price of industrial land in China *Research of Institutional Economics* **10**(1) 184-99 *(in Chinese)*

[3] Xu Y 2009 The land price formation mechanism of Beijing industrial parks *Commercial Research* **20**(1) 57-61 *(in Chinese)*

[4] Yang L 2012 The reasons for the governmental failure in the allocation of industrial land in China *China Land Sciences* **26**(8) 36-41 *(in Chinese)*

[5] Wang J 2012 Industrial agglomeration, government intervention and industrial land prices: an empirical study on 35 cities in China *China Land Sciences* **26**(9) 12-20 *(in Chinese)*

[6] Liu J 2013 The analysis of the influencing factors of urban industrial land prices in China: an empirical study based on panel data of 35 cities *Price: Theory &Practice* **21**(7) 56-7 *(in Chinese)*

[7] Bicher M 2000 An experimental analysis of multi-attribute auction *Decision Support Systems* **29**(2) 249-68

[8] Che Y K 1993 Design competition through multidimensional auctions *RAND Journal of Economics* **24**(4) 668-80

[9] Branco F 1997 The design of multidimensional auctions *Journal of economics* **28**(1) 63-81

[10] Parkes D C, Kalagnanam J 2004 Iterative multi-attribute vickrey auctions *Management Science* **9**(6) 11-17

[11] David E, Azoulay-Schwartz R, Kraus S 2002 Protocols and strategies for automated multi-attributes auctions *Proceedings of the 1st Conference on Autonomous Agents and Multi-agent Systems* Italy Bologna 77-85

[12] David E, Azoulay-Schwartz R, Kraus S 2002 An English auction protocol for multi-attributes items *Revised papers from the workshop on Agent mediated electronic commerce: IV Designing Mechanisms and Systems* **2531** 52-68

[13] David E, Azoulay-Schwartz R, Kraus S 2006 Bidding in sealed-bid and English multi-attribute auctions *Decision Support Systems* **42**(3) 527-56

[14] Zeng X 2010 Research on model and mechanism of multi-attribute auction based on game theory *Master thesis* Harbin Institute of Technology 35-45

Authors

Jian-gang Shi, born in 1962, Shanghai, P.R. China

Current position, grades: professor at the School of Economics and Management, Tongji University, China.
University studies: B.Sc. in Surveying Engineering at Tongji University in China. M.Sc. and PhD at Tongji University in China.
Scientific interests: urban development and management, computer simulation, project management, evaluation of real estate.
Publications: more than 60 papers.
Experience: teaching experience of 28 years, 30 scientific research projects.

Jin-can Liu, born in April, 1977, Qi County, Henan Province, P.R. China

Current position, grades: PhD candidate at the School of Economics and Management, Tongji University, China.
University studies: Master degree in project management at Tsinghua University in China in 2009.
Scientific interests: computer simulation, urban development and management.
Publications: more than 5 papers.
Experience: teaching experience of 4 years, 3 scientific research projects.

Statistical analysis and prediction of Qingdao urban consumer price index

Linlin Dai*

School of Qingdao Huanghai College, No.3111, Ling Hai Road, Qingdao Shandong Province, China

Abstract

In recent years, Qingdao faces inflationary pressure along with the fast development of economy. The urban consumer price index (CPI) is an important measure of regional inflation level. Thus, analysing the influencing factors of the CIP and determining the key factors is necessary to supply scientific evidence for the economic management departments to adjust the industrial structure. This paper filtrates and eliminates six indexes likely to affect urban CPI of Qingdao based on Studentized residual, correlation analysis, and multiple collinearity diagnosis. This paper also performs a statistical analysis of the urban CPI of Qingdao from 2005 to 2012 and draws the conclusion that food price is the largest index that affects CPI. Basing on this conclusion, corresponding prediction models are established to supply statistics for departments that adjust the economy for the future.

Keywords: Studentized residual, correlation analysis, collinearity diagnosis, multiple linear regression model

1 Introduction

Along with the rapid development of the Chinese economy, the consumption level is rising and has become an important measure of a regional economy. Consumer price index (CPI) has also become an important measure of the level of inflation. Much research has been conducted on CPI. For example, Zhu Wei and Zhong Weijian have researched the application of the ARMA model in CPI [1]. He Weiwei and Tianhao have analysed the influencing factors of CPI according to the VAR model [2]. Yu Hongyan and Chu Delin have analysed the dynamic correlation between the Chinese customer price index and food price index according to the H-P filtering method [3]. Currently, although the economy of Qingdao is developing rapidly, it also suffers from inflationary pressure. The major reasons for this pressure are that industrial structure is irrational to some extent, and some industries fail to meet the needs of the society. This paper, based on an analysis of Qingdao urban CPI statistics, attempts to determine the positive and negative factors that affect economic development and offers some proposals about optimizing the industrial structure to reduce the pressure of inflation.

The possible factors affecting urban CPI (X1) include the following: the commodity retail price index (X2), the food price index (X3), industrial producers purchase price index (purchasing price index of raw material, fuel, and power) (X4), the producer price index (X5), the price index of investment in fixed assets (X6), and service price index (X7). The index data are collected from the online Statistical Yearbook of Qingdao from 2005 to 2012 and other related information sources. After preparing and analysing the data in advance, the Studentized residuals and related analysis are used to determine the degree of correlation among these indexes and to further filter out the most important indicator of the urban CPI. Multi-collinearity diagnosis is employed for the selected variables to estimate the capability of the multiple regression model. Finally, the selected indicators are used to establish a regression model for the principle component of the urban CPI and to further analyse the effectiveness and accuracy of the model. Given that the real effect factor is fully considered, this model reflects the status of multiple layers in CPI effect well, and both its fitness and significance are superior.

An applied technical route is used as follows: Raw data extraction - gross error analysis (Studentized residual analysis) - related analysis, - multicollinearity diagnosis - multiple regression model is established.

2 Establish mathematical model and solve model

2.1 DATA PREPROCESSING

2.1.1 Raw data extraction

Given that a human extracted the raw data used, some measurement errors are inevitable. Thus, the data should be pre-processed first. The index data of some years are incomplete. Thus, we finally chose Qingdao data from 2005 to 2012 to analyse the CPI of this city. Although the sample size decreased, the actual test data can be reflected maximally. The primary index data [4] are shown in Table 1:

Corresponding author e-mail: dailinlin79@163.com

Year	X1	X2	X3	X4	X5	X6	X7
2005	102.3	99.3	102.1	106.5	101.4	103.1	101.4
2006	100.9	99.7	101.7	105.8	101.1	102.6	100.8
2007	104.5	102.7	111.6	106.7	101.5	104.1	102.4
2008	104.7	103.9	111.7	115.9	105.3	111.0	99.8
2009	100.5	98.6	101.6	90.4	95.8	94.6	100.6
2010	102.2	101.4	106.4	112.2	103.8	104.7	100.4
2011	105.0	104.5	111.1	109.4	104.9	107.0	103.1
2012	102.7	101.7	104.3	97.0	98.6	100.0	102.1

2.1.2 Gross Error Analysis

Bulky errors or gross errors, also known as outliers, are individual values in the samples that obviously deviate from the other samples [5]. When processing the data, if the outliners mixed in the measurement data are not eliminated, the statistical analysis would be less accurate. Conversely, if normal data have been eliminated as outliers, the statistical analysis would have less information. Hence, correct result must first determine and process the outliners rationally.

This paper uses the Studentized residual to analyse the gross error of the original data. The basic concepts of Studentized residuals [6] are depicted as follows:
Assuming n times independent measurements are obtained, the column measurement X is obtained as follows:

$$X = \{x_1, x_2, ..., x_i, ..., x_n\} \ (i = 1,2,...,n) .\tag{1}$$

The mean $\bar{x}$ and residual v_i respectively are:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} ,\tag{2}$$

$$v_i = x_i - \bar{x} .\tag{3}$$

Definition 1: Sample Studentized residual is the ratio of residuals to the sample standard deviation, that is:

$$y_i = \frac{v_i}{s} .\tag{4}$$

In the following formula, s stands for a sample standard deviation, and the value of s is given by:

$$s = \sqrt{\frac{\sum_{i=1}^{n} v_i^2}{n}} .\tag{5}$$

Definition 2: The absolute value of the sample Studentized residual is the ratio of the absolute value of the residuals to the absolute value of the sample standard deviation, that is

$$|y_i| = \frac{|v_i|}{s} .\tag{6}$$

"Studentized" here refers mainly to the common residual without homogeneity of variance. We suppose that expectation is 0 and variance is 1 [7] for easier comparison.

According to the above definition of Studentized residual, the annual data of each indicator are processed. The result is shown in Figure 1. When the absolute value of the Studentized residual is more than 3, the data should be eliminated as variable data, that is, the gross error. However, Figure 1 shows that each absolute value of the indexes' Studentized residuals is less than 3. Thus, the above index data have no gross error and can be directly used for further analysis.

The above value indexes are all prices that share the same unit. Thus, they do not need to be further standardized.
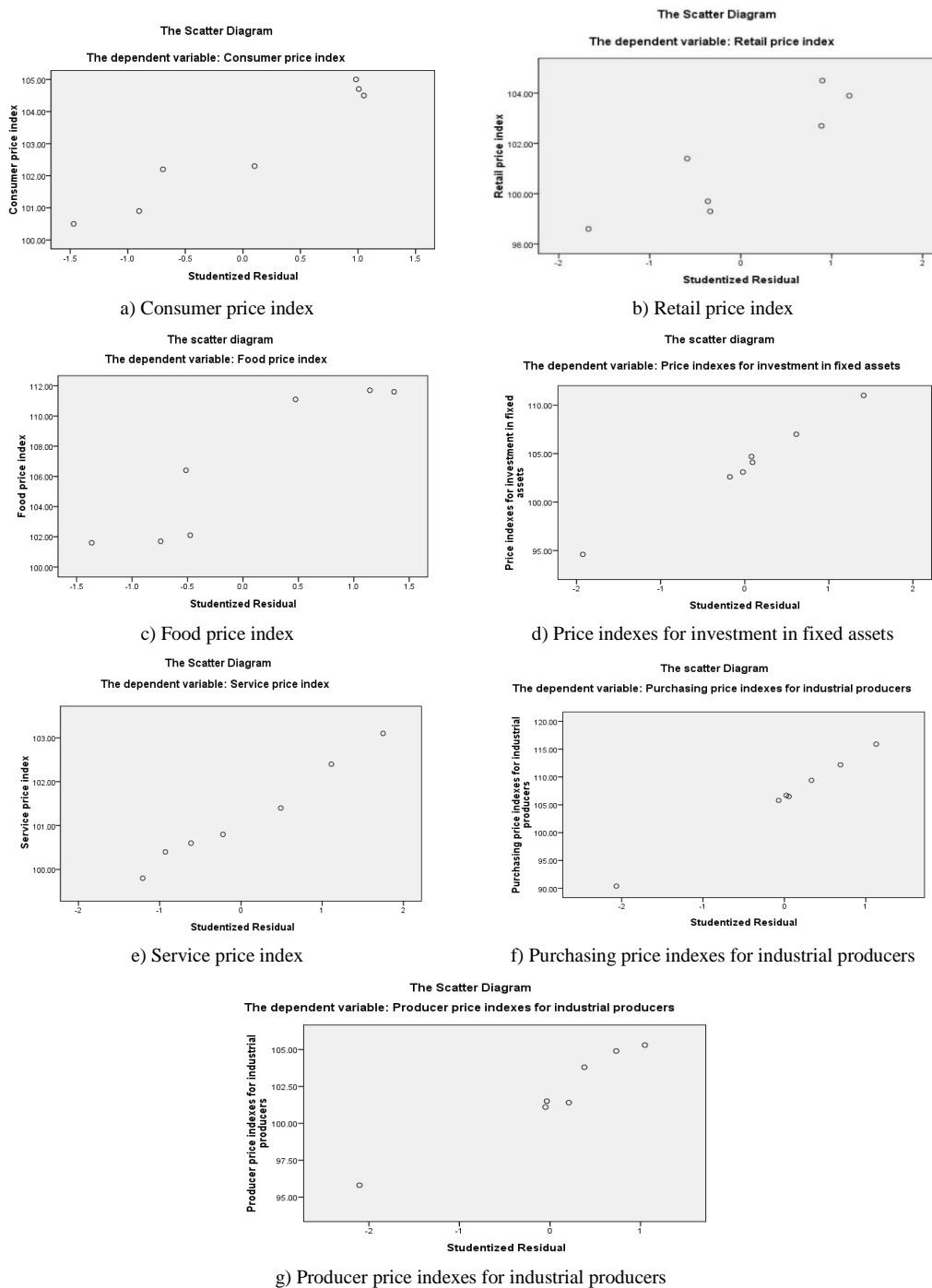
a) Consumer price index

b) Retail price index

c) Food price index

d) Price indexes for investment in fixed assets

e) Service price index

f) Purchasing price indexes for industrial producers

g) Producer price indexes for industrial producers
FIGURE 1 Studentized residual figure of each index variable data

2.2 INDEX VARIABLES ANALYSIS BASED ON CORRELATION ANALYSIS

Correlation analysis is a statistical method that studies the quality of relationships between variables. Through correlation analysis, this paper presents a preliminary understanding between urban CPI and other indexes. Thus, factors that are not significant are eliminated to reduce model dimension.

Pearson correlation coefficient between any two variables can be calculated according to Equation (7):

$$\rho = \frac{\sum_{k=1}^{n}(X_{ik} - \overline{X}_i)(X_{jk} - \overline{X}_j)}{\sqrt{\sum_{k=1}^{n}(X_{ik} - \overline{X}_i)^2 \sum_{k=1}^{n}(X_{jk} - \overline{X}_j)^2}}, \qquad (7)$$

$\overline{X}$ is the average of single variables, and X_{ik} represents the *k-th* variable in the *i-th* data.

SPSS software was used for the correlation analysis of the data in Table 1. The Pearson correlations among each index value can be obtained, as shown in Table 2:

TABLE 2 Pearson correlation coefficient matrix among each index

		X1	X2	X3	X4	X5	X6	X7
X1	Pearson correlation	1	.936**	.938**	.773*	.436	.614	.709*
	Significant (both sides)		.001	.001	.024	.280	.105	.049
X2	Pearson correlation	.936**	1	.935**	.777*	.357	.618	.740*
	Significant (both sides)	.001		.001	.023	.386	.102	.036
X3	Pearson correlation	.938**	.935**	1	.750*	.274	.628	.702
	Significant (both sides)	.001	.001		.032	.511	.095	.052
X4	Pearson correlation	.773*	.777*	.750*	1	−.040	.953**	.963**
	Significant (both sides)	.024	.023	.032		.926	.000	.000
X5	Pearson correlation	.436	.357	.274	−.040	1	−.148	.008
	Significant (both sides)	.280	.386	.511	.926		.727	.985
X6	Pearson correlation	.614	.618	.628	.953**	−.148	1	.968**
	Significant (both sides)	.105	.102	.095	.000	.727		.000
X7	Pearson correlation	.709*	.740*	.702	.963**	.008	.968**	1
	Significant (both sides)	.049	.036	.052	.000	.985	.000	

**. At the .01 level (both sides) significant correlation
*. At the 0.05 level (both sides) significant correlation

Generally speaking, two vectors will have strong correlation when the absolute value of the Pearson correlation coefficient is greater than 0.8. According to this criterion, the correlation is not strong between urban CPI and investment in fixed assets, service price, industrial producers purchase price, or producer price. Therefore, these four indexes can be eliminated first. Then, the rest of the indicators are analysed.

2.3 MULTIPLE COLLINEARITY DIAGNOSIS

Two variables remain after eliminating unimportant variables, and multiple collinearity diagnosis must be used to determine whether a linear correlation is present between these two variables.

Collinearity can be measured through tolerance [8]. If the value is less than 0.1, then a collinearity problem may be present between the two indexes. The tolerance between retail price and food price in SPSS software is shown in Table 3:

TABLE 3 Tolerance calculation table

Dependent	Independent	Collinearity statistics	
		Tolerance	VIF
X2	X3	1.000	1.000
X3	X2	1.000	1.000

Table 3 shows that the tolerance is close to 1, and the variance inflation factor is also close to 1. Thus, the independent variables X2 and X3 have a weak collinearity problem.

2.4 URBAN CPI BASED ON MULTIPLE LINEAR REGRESSION MODEL

A multiple linear regression model is a linear regression model with several explanatory variables. Such models are used to illustrate the linear relationship among explained variables and several explanatory variables, and its regression equation is given by:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n . \qquad (8)$$

In this paper, the multiple linear regression model is analysed using SPSS software, and its process is as follows:

On the basis of previous correlation analysis, two indicators are significant: the commodity retail price index (X2) and the food price index (X3). The original data of these two indexes in Table 1 are extracted and are used to establish multiple linear regression models by using the stepwise regression method. The results are summarized in Tables 4-7:

TABLE 4 Input/remove variables

Model	Input variable	Remove variable	Methods
1	X3	—	Step by step (guidelines: The probability of F-to-enter <= .050. The probability of F-to-remove >= .100).

a. dependent: X1

TABLE 5 Model summary

Model	R	R2	Adjusted R2	Standard estimate error
1	.938a	.880	.861	.64430

a. Independent: (constant), X3

TABLE 6 Coefficient table

Model		Standardized coefficients		Standard coefficient	t	Sig.
		B	Standard error	Trial version		
1	(constant)	65.056	5.689	—	11.435	.000
	X3	.356	.053	.938	6.648	.001

a. Dependent: X1

TABLE 7 Variables have been ruled out

Model		Beta In	T	Sig.	Partial correlation	Collinearity statistics
						tolerance
1	X2	.464a	1.210	.280	.476	.126

a. Prediction of the variables in the model: (constant), X3
b. Dependent: X1

In the process of stepwise regression, the variable X2 has been removed. Therefore, only X3, the food price index, has been left as an independent variable. Table 5 shows that the model has extremely good fitting. Table 6

shows that the variable coefficient is significant ($\rho < 0.05$). According to the above regression results, the prediction model of urban CPI can be concluded as:

$$X_1 = 65.056 + 0.356X_3 + \varepsilon . \qquad (9)$$

Figure 2 illustrates that the observed and predicted values tend to agree with each other, which further indicates the good quality of the model.
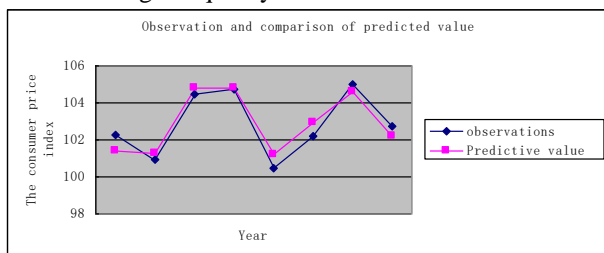

FIGURE 2 Comparison chart of observed and predicted CPI values

In this paper, the statistical analysis and prediction of the Qingdao urban residents' consumption index have been proposed. The close relationship between the urban CPI and food prices will greatly help relevant government departments make correct decisions. Government departments must strictly control food prices to ensure the stability of the urban CPI, which will maintain the balance of the industrial economy in Qingdao City as well as maintain a healthily developed and sustained economy.

3 Conclusion

This article performs statistical analyses and predictions of influencing factors on urban CPI in Qingdao and concludes that a close relationship exists between urban CPI and food price. Figure 2 illustrates that a stable long-run equilibrium relationship exists between CPI and food price index. Over the past decade, food price has been the main influencing factor of CPI. Additionally, through Studentized residual and correlation analysis, we can infer that non-food prices also affect CPI to some extent, and the effect will gradually increase with the development of the society. Therefore, the government department concerned must focus on the effect of food prices when considering CPI and price level. Meanwhile, in the long process of revising food prices to stabilize CPI and price level, other relative prices, especially some affecting factors that change widely or greatly, must not be ignored. Doing so will promote the sound, balanced, and sustained development of the industrial economy.

Acknowledgments

References

[1] Zhu W, Zhong W J 2008 Application of ARMA Model in Predicting the Consumer Price Index *Finance & Economy* **16** 82-3 *(in Chinese)*
[2] He W W, Tian H 2007 Analysis of Influencing Factors of Consumer Price Index *Statistics and Decision* **23** 66-8 *(in Chinese)*
[3] Yu H Y, Chu D Y 2011 Analysis of The dynamic Correlation between China's Consumer Price Index and Food Price *Index Price: Theory & Practice* **2** 52-3
[4] Qingdao Statistical Information Network *statistical yearbook of Qingdao*. http://www.stats-qd.gov.cn/statsqd/Columns/2013nj.asp?typeid=1930&videos=&jms=230/ 29 Jan 2013 *(in Chinese)*
[5] Li Q K, Lv Z P 2008 Fuzzy Adaptive Kaiman Filtering Algorithm Based on the Statistics of *t* Distribution *Acta Geodaetica et Cartographica Sinica* **37**(4) 228-9 *(in Chinese)*
[6] Liu X L, Li Q H 2008 The Discussion about Robust Estimation based on The Equivalent Weight Function constructed by Studentized Residual *Modern Surveying and Mapping* **31**(6) 207-8
[7] Wang Z Y, Zhang H B, Liu Z M 2006 The Novel Method for Outliers' Rejection of the Studentized Residual Error *Chinese Journal of Scientific Instrument* **27**(6) 624-8 *(in Chinese)*
[8] Lin L, Ding W X, Brower D L 2012 Collinearity alignment of probe beams in a laser-based Faraday effect diagnostic *Review of Scientific Instruments* **83**(10) 320-3

Author

Linlin Dai, born in August, 1979, Liaocheng District, Shandong Province, P.R. China

Current position, grades: lecturer at the School of finance and economics Qingdao Huanghai college, China.
University studies: Master Degree in education at Guizhou Normal University in China. Bachelor Degree in science at Liaocheng University in China.
Scientific interests: Economics statistics.
Publications: 6 papers.
Experience: teaching experience of 8 years, 2 scientific research projects.

Operation Research and Decision Making

Reliability and safety verification of the new collision avoidance strategy for Chinese train control system

Junting Lin[1]*, Xiaoming Wang[1], Jianwu Dang[1]

[1] *School of Automation & Electrical Engineering, Lanzhou Jiaotong University, Anning West Road, Lanzhou, China*

Abstract

Although equipped with comprehensive and complex technologies on the railway to avoid collisions between trains, such as train control system extensively deployed in the infrastructure, there are still enormous amount of collision between trains. The safety of railway operation mainly depends on the correct operation of the interlocking and train control system, because only the control centre has an overall overview of the traffic situation, and a train driver could only be warned of a possible collision if the operation centre decides it. Experiences from the aviation, the maritime navigation and road transportation have shown that the probability of collisions could be significantly reduced with infrastructure-less collision avoidance system. In this paper, an enhanced safety strategy, namely, Collision Avoidance System overlay Train Control System (CASOTCS) for Chinese railway was provided, which was based on direct vehicle-to-vehicle communication and independent of the regular control mechanism. CASOTCS can receive and evaluate the information broadcasted by other neighbouring trains all the time, which will trigger collision alert and provide a solution for the collision when a potential collision is detected. The unit architecture and its key issues of CASOTCS were also discussed. After the analysis of reliability and safety of both the traditional train control system and the enhanced CASOTCS, the reliability block diagram and isomorphic Markov model were established in the paper. Through the calculation of the indexes of the reliability and safety about the two kinds of control systems, a significant conclusion has been made, that is, the reliability and safety of the train control system plus collision avoidance system are much higher than that of the traditional one and CASOTCS has the ability to increase the reliability, safety and efficiency in the future.

Keywords: Collision Avoidance, Chinese Train Control System, Safety Verification, Reliability, Markov

1 Introduction

Actual statistics of the European Railway Agency (ERA) [1] and German Aerospace Centre (DLR) show that there are serious train-collision accidents all over the world, despite millions of money have been invested in infrastructure equipment. In order to increase safety in railway system, many countries are installing Train Control Systems (TCS), mainly centralized management ones, where trains are monitored by control centre.

Additionally, TCS used in rail transportation is heavily infrastructure-based, it is therefore clear that there is a need for a completely new safety system or overlay system that improves safety while reducing infrastructure and maintenance costs[2].

Experiences from the state of the art of collision avoidance system for different transport means, i.e. Traffic Alert and Collision Avoidance System (TCAS) [3], Automatic Dependent Surveillance Broadcast (ADS-B) [4], Automatic Identification System (AIS) [5] and Car-to-Car (C2C) [6] have shown us that the probability of collisions can be significantly reduced collision avoidance system, which do hardly require infrastructure components To meet the requirements of the systems for high safety and high reliability, various redundancy and restructuring cells are widely applied in electronic product design to improve the reliability and safety of the systems, as well as online fault diagnosis technology. In terms of safety, reliability and cost, dual hot spare dynamic redundancy structure is a kind of ideal design scheme, and has already been widely applied in modern railway signal systems. Analysis of reliability and safety of redundant systems have been researched extensively. As in [7], He applies fault tree to investigate the safety of the switch control unit of all-electronic computer interlocking system. Zhang [8] analyzes and compares the reliability and safety of several common-used redundant structures. Hence, the paper aims at dynamic redundant communication systems between trains, which has already been widely applied in Chinese high-speed railway, and established reliability block diagram of centralized TCS and CASOTCS. Synthetically considering the influence of many factors such as coverage of diagnostic systems, online maintainability and many failure modes, we finally established the Markov model and analyzed the safety indexes. The result shows collision avoidance system has the ability to increase safety and efficiency.

The rest of this paper is organized as follows: Section 2 we describe the state of art in Chinese train control system. Section 3 introduces the collision avoidance systems in maritime, air, and road transport, and

---

* *Corresponding author* e-mail: linjt@mail.lzjtu.cn

discusses the unit structure of Collision Avoidance System overlay Train Control System (CASOTCS) and its key issues. Section 4 verifies the reliability and safety of CASOTCS. Finally, conclusion is given in section 5.

2 State of the art in CTCS

Conventional trains in China were manually controlled and operated by drivers based on trackside interlocking and blocking, in conjunction with various train signaling and surveillance devices [9]. As the high-speed railway systems being rapidly developed in China recently, the old Track Based Train Control (TBTC) has gradually been replaced by the modern Communication Based Train Control (CBTC).

The Chinese Train Control System (CTCS) is a train control system implemented since 2002, which is specified for compliance with the high-speed and the conventional interoperability and directives. CTCS by nature is an automatic train protection (ATP) system, based on cab-signaling and signal aspects as well as continuous tracking to the data transmission on the train system. The movement authority (MA) and the corresponding line information are transmitted to the control unit of the train and then being displayed in the cab for the driver as commands or references. A train with complete CTCS equipment and functionality can operate on any CTCS line without much technical restriction.

Based on the signaling system of Chinese railway network, Chinese Train Control System (CTCS) has been divided into five levels, namely CTCS level-0, 1, 1, 2, 3 and 4, referring to European Train Control System (ETCS) [10]. The definition of the levels depends on how the line is equipped and the way the information is transmitted to the train. Interoperability is necessary for the train control system to achieve joint operation among Dedicated Passenger Lines (DPL, another name for high-speed railway in China) and rebuilt lines, where CTCS levels-2, 3 and 4 are back-compatible with lower levels.

CTCS level 0. It consists of the existing track circuits, universal cab signaling (the digital, microprocessors-based cab signaling that be compatible with the six kinds of track circuits on Chinese Railway Network, designed by the research team of Beijing Jiaotong University ten years ago) and train operation supervision system. With level 0, wayside signals are the main signals and cab signals are the auxiliary signals. It is the most basic mode for CTCS. It is no necessary to upgrade the wayside systems for CTCS level 0. The only way to realize the level 0 is to equip with the on-board system. CTCS level 0 is only for the trains with the speed less than 120km/h.

CTCS level 1. It consists of the existing track circuits, transponders (or Balises) and ATP system. It is for the train with the speed between 120km/h and 160km/h. For this level, the block signals could be removed and train operation is based on the on-board system, ATP, which is called as the main signals. Transponders (Balises) must

be installed on the line. The requirements for track circuit in blocks and at stations are higher than that in the level 0. The control mode for ATP could be the distance to go or speed steps.

CTCS level 2. It consists of digital track circuits (or analog track circuits with multi-information), transponders (Balises) and ATP system. It is used for the trains with the speed higher than 160 km/h. There is no wayside signaling in block for the level 2 anymore. The control mode for ATP is the distance to go. The digital track circuit can transmit more information than analog track circuit. ATP system can get all the necessary information for train control. With this level, fixed block mode is still applied. The system indicates the special feature of Chinese railway signaling.

CTCS level 3. It consists of track circuits, transponders (Balises) and ATP with GSM-R. In the level 3, the function of the track circuit is only for train occupation and train integrity checking. Track circuits no longer transmit information concerning train operation. All the data concerning train operation information is transmitted by GSM-R. GSM-R is the core of the level. At this level, the philosophy of fixed block system is still applied. It was firstly used on the Wuhan-Guangzhou High-speed Railway services, where trains have speeds up to 350 km/h on DPL, which was started operation in December 2009. It has two subsystems: ground subsystem and onboard subsystem. The ground subsystem includes Balises, track circuits, a wireless communication network (GSM-R), and a Radio Block Centers (RBCs). The onboard subsystem, on the other hand, includes onboard devices and an onboard wireless module.

CTCS level 4. It is the highest level for CTCS. Moving block system function can be realized by the level 4. The information transmission between trains and wayside devices is made by GSM-R. GPS or transponders (Balises) are used for train position. Train integrity checking is carried out by on-board system. Track circuits are only used at stations. The amount of wayside system is reduced to the minimum in order to reduce the maintenance cost of the system. Train dispatching can be made to be very flexible for the different density of train operation on the same line.

From the above categories, we can see that the safety of railway operation is mainly ensured by the centralized infrastructures, such as interlocking and block system, which set and lock the train route. The trains are equipped with the onboard part of the train control system, which triggers automatically a braking if the train passes a signal at danger.

The safety of railway operation depends on the correct operation of the train control system where the trains are monitored by devices located along the rail. These devices send the collected information to an operation center named Train Control Center (TCC) or Radio Block Center (RBC) that can pass specific instructions to the train, as shown in Figure.1. Human

Lin Junting, Wang Xiaoming, Dang Jianwu

errors and the communication failure between train and control center lead to most of the accidents. Additionally, only the operation center has an overall overview of the traffic situation, and a train driver could only be warned of a hypothetical collision if the operation center decides it.
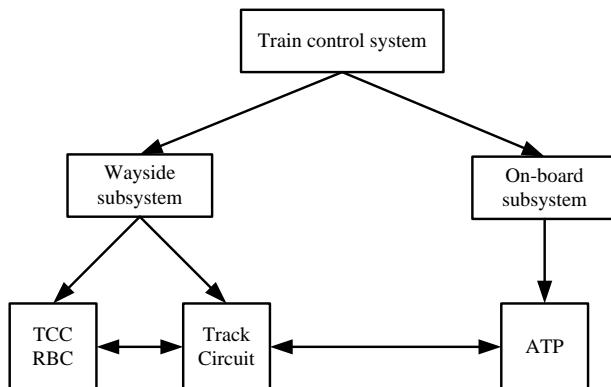

FIGURE 1 Control strategy of CTCS

3 State of the research in CAS and applicability to CTCS

3.1 CAS AND ITS APPLICABILITY

Collision avoidance systems related with aviation, maritime, and road transport systems have gone through a rapid evolution and improvement with significant growths.

Aviation: TCAS relies on the Secondary Surveillance Radar (SSR). By means of this radar, a TCAS equipped aircraft interrogates other aircraft in its vicinity and listens for the transponder replies [11]. Computer analysis of these replies determines which aircraft represent potential collision threats. Consequently Traffic Advisories (TA), in TCAS I, and resolution advisories (RA), in TCAS II and III, can be provided. ADS-B conform the evolution of TCAS. Based on the Global Positioning System (GPS), an aircraft can automatically broadcast its identification address. GPS derived latitude and longitude, altitude and the 3D velocity.

Maritime navigation: The maritime surveillance application of AIS consists of a continuous interchange of driving data from ships like GNSS position, speed and direction, as well as relevant information like identification numbers, length and beam, ships draught, route plan etc..

Road Transport: The Car 2 Car Communication Consortium aims to establish an open European industry standard for the ad-hoc communication between vehicles and vehicles to infrastructure, which is currently in the development phase. Together with the position determined by a GNSS receiver this information is broadcast to the vehicles in the vicinity which may detect the presence of traffic jam by exploiting the received information.

Applicability to Chinese Railway: While maritime, air, and road transport have a vehicle integrated collision

avoidance system available or in the development phase, we find no satisfactory solution of this type of technology in railway transportation. Therefore, it is necessary to develop a system that will allow the train conductors to have an up-to-date accurate knowledge of the traffic situation in the vicinity and act in consequence. The system is intended to not rely on components in the infrastructure, this way substantially reducing its rollout and maintenance costs, as well as inherently providing a migration strategy. The basic idea is to communicate relevant own context information to all other nearby trains.

Experiences from the TCAS and ADS-B, as well as AIS and C2C, have shown that the probability of collisions can be significantly reduced with collision avoidance systems. In this section, we introduce our CASOTCS approach consisting only of Train-to-Train communication components, i.e. without the necessity of extensions of the railway trackside infrastructure, as depicted in Figure 2. Each train determines its position, direction and speed using Global Navigation Satellite System (GNSS) and broadcasts this information, complemented with other important information in the region of its current location. This information can be received and evaluated by other trains, which may – if a potential collision is detected – lead to traffic alerts and resolution advisories up to direct interventions (usually applying the brakes).

3.2 KEY ISSUES OF CASOTCS

In the designed CASOTCS, each trains has to calculate its own position and movement vector and broadcast this information as well as additional data like vehicle dimensions to all other trains in the area. Thus, the driver's cabin could be equipped with a MMI showing the position of the other vehicles in the region. CAS vital computer analyzes all received context information, the own position and movement vector and an electronic track map detects possible collisions, displaying an alert signal, and advising the driver of the most convenient strategy to follow in order to avoid the danger. The system can take into account different danger sources, like advancing trains or road vehicles or obstacles, and classify them according to a specific scale.

In principle, the CASOTCS is very similar to TCAS/ADS-B used as a "safety overlay system" in aeronautical transport, which is as well controlled by a number of operation centers. The advantages of such infrastructure-less collision avoidance overlay system are:
- Collision avoidance system works independently of the regular traffic control mechanism.
- No changes are required on the existing train control system infrastructure.
- Direct train-to-train communication as the supplement of GSM-R to enhance safety.
- Continuous "roll-out" is possible.

Lin Junting, Wang Xiaoming, Dang Jianwu

- Lower costs compared to infrastructure based system.
- Potential to increase efficiency, e.g. shorter distances between trains.

The railway specific boundary conditions require new design considerations [12]. The major components of an infrastructure-less collision avoidance system for CTCS are position detection, direct train-to-train communication and collision surveillance resolution.

3.2.1 Speed detection and positioning system

The precise detection of the position plays a key role for collision avoidance system and train control system. The positioning system proposed in CASOTCS determines the position of the train independently from the trackside equipment. Because of inadequate fulfillment for safety-critical application, GNSS-based positioning system has to be combined with at least one additional positioning technique such as digital map, odometer or inertial system.

3.2.2 Direct train-to-train communication

Simulations have been done in shown that using an air brake system to stop a train at initial speed of 300 km/h, it will take 4.1 km for the train to arrive at a complete stop. In order to design a suitable direct train-to-train communication system, a six phase work approach was conducted: preliminary analysis and selection of an adequate frequency band, characterization of the propagation channel, MAC layer design, physical layer design and verification of the system [13, 14]. The following challenges arise designing the system: no infrastructure can be used, transmission is broadcast and the system will run in high speed line.

3.2.3 Collision surveillance resolution

Each collision avoidance system unit produces messages with a fixed length that are broadcasted with a variable rate based on its own speed and the vicinity traffic situation. The Train Number (TN) or the Locomotive Number (LN), the current speed, the braking distance, and the forward and the backward length of the train with respect to the localization unit are included in the messages to allow other vehicles to identify potential collision (head-on, rear-end or flank collision). To avoid collision, a Collision Alert (CA) signal as the first step shall warn the driver in case of a detected close approach to another CAS unit on a collision course, and the driver is prepared to perform a braking operation in the second step. Reliability of collision detection in time is the most important property in CAS, and it is the estimated braking distances that can distinguish collision scenarios from regular operation.
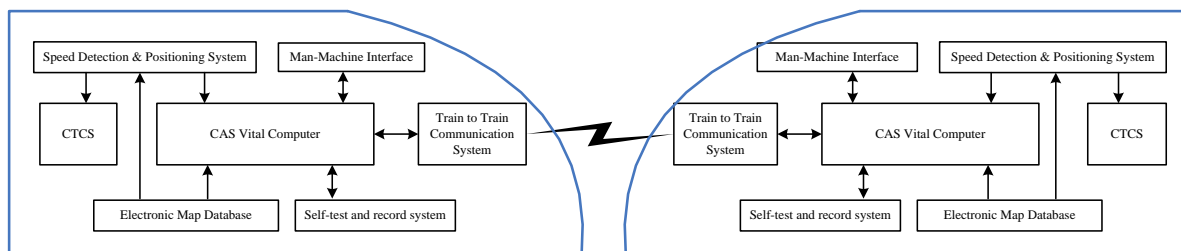
FIGURE 2 Unit architecture CASOTCS

4 Reliability and safety verification of CASOTCS

Reliability is defined as the probability that a system (component) can complete the regulated function under specified conditions and in range of the prescribed time. To express this relationship mathematically we define the continuous random variable T to be the time to failure of the system (component) [15]. Then reliability can be expressed as

$$R(t) = \begin{cases} P(T > t) & t \geq 0 \\ 0 & t < 0 \end{cases},$$ (1)

where $R(t) \geq 0$, $R(0) = 1$ and $\lim_{t \to \infty} R(t) = 0$. In analyzing a complex system, an alternative approach is to determine an appropriate reliability or reliability model for each component of the system, and by applying the rules of probability according to the configuration of the components within the system, compute a system's reliability.

Corresponding to the reliability, the unreliability can be defined as

$$F(t) = P(0 \leq T \leq 0) = 1 - R(t)$$ (2)

Thus, mean time to failure (MTTF) can be presented by

$$MTTF = E(T) = -\int_0^{+\infty} td[R(t)] = \int_0^{+\infty} R(t)dt.$$ (3)

Safety refers to the ability that the system could not generate the dangerous side outputs when the fault occurs. Reliability, availability, and *MTTF* are only for the normal work of the system concerned. As the system

enters into the failure state from the normal one, we can say that it breaks down and terminates the job, and cannot continue to perform the regulated function. At the moment, there are two significant states to need to be considered, that is, the safety failure state and the dangerous failure state. The former is corresponding to a safe failure probability (PFS) and the latter is corresponding to a dangerous failure probability (PFD) .

Thus, the unreliability may be written as

$$F(t) = PFS(t) + PFD(t) .\qquad(4)$$

In terms of the repairable systems, the availability is

$$A(t) = 1 - [PFS(t) + PFD(t)] .\qquad(5)$$

The safety availability is different with the availability, and defined as

$$S(t) = 1 - PFD(t) .\qquad(6)$$

Through online condition monitoring and fault diagnosis, we can reduce maintenance time and control the implementation of some tolerant structure. The diagnostic coverage rate can be applied to express the power of the diagnostic system, which reflects the probability that if a failure occurs it can be detected. In the numerical value, it equals the sum of the detected failure rates is divided by total failure rate. Hence, it is necessary to consider the influence of failure detection system when analyzing a system safety [16].

Without consideration common cause, the failure rate of the channel is partitioned as the two parts, that is, the safety failure rate λ^S , and the failure rate λ^D , and then

$$\lambda = \lambda^S + \lambda^D .\qquad(7)$$

Consider the role of online diagnostic systems, the failure rate λ^S is divided as the two parts, that is to say, the detected safety failure rate λ^{SD} ,and the undetected safety failure rate λ^{SU} , and then

$$\lambda^S = \lambda^{SD} + \lambda^{SU} .\qquad(8)$$

For λ^D , likewise, we let λ^{DD} expresses the detected dangerous failure rate, and λ^{DU} be the undetected dangerous failure rate. We have

$$\lambda^D = \lambda^{DD} + \lambda^{DU} .\qquad(9)$$

Components within a system may be related to one another in two primary ways: in either a serial or a parallel configuration. In series, all components must function for the system to function, while in a parallel, at least one component must function for the system to function.

To simplify the notation, the argument of $R(t)$ will be dropped [17] In this case, it is understood that all reliabilities are to be evaluated for the same point in time

t . Generalizing to n mutually independent components, serial system reliability is expressed as

$$Rss = \prod_{i=1}^{n} R_i ,\qquad(10)$$

where R_i represent the reliability of the unit i . System reliability for n parallel and independent components is given by

$$R_{SN} = 1 - \prod_{i=1}^{n} (1 - R_i).\qquad(11)$$

Systems typically contain components in both serial and parallel relationships. To compute the system reliability, the network may be broken into serial or parallel subsystems.

There are dynamic redundant structures applied in modern train control system, which are here respectively defined as the fundamental mode is shown in Figure 3. The enhanced mode including train-to-train communication channel is shown in Figure 4.
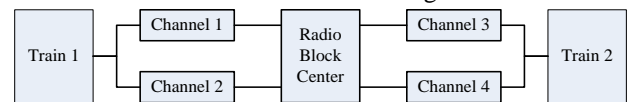

FIGURE 3 Fundamental mode of train to train communication channel

$$R_{S1} = \left[1 - (1 - R_{Ch1})(1 - R_{Ch2})\right] \times$$
$$R_{RBC} \times \left[1 - (1 - R_{Ch3})(1 - R_{Ch4})\right] ,\qquad(12)$$

where R_{Ch1} , R_{Ch2} , R_{Ch3} and R_{Ch4} represent the reliability of Channel1, 2, 3 and 4 respectively, and R_{RBC} is the reliability of RBC.
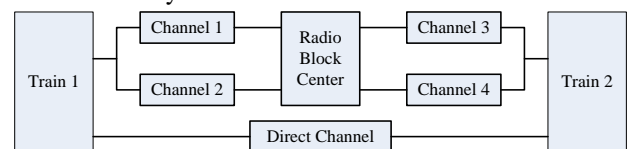

FIGURE 4 Enhanced mode of communication channel in CASOTCS

Let the direct channel reliability be R_{DC} , we have

$$R_{S2} = 1 - (1 - R_{S1})(1 - R_{DC}) .\qquad(13)$$

According to equation (12) and equation (13), the diverse reliability and their difference are calculated as shown in Table 1 and Figure 5. Clearly, system reliability relies on the reliability of units. The higher is the unit reliability, the higher is the system reliability. Enhanced mode is more reliable than the fundamental mode, but differences between them are minimized.

The security of single channel with single version software only relies on channel own security, which is difficult to meet the safety requirements. Dual channel with double version software cannot only detect the most software faults, but also can find the hardware faults, and the requirements on the hardware are not so high. If the self-diagnosis program detects and prompts the

emergence of a failure, then the failure can be immediately repaired, and otherwise it may not still known. To be able to find the failures early, the regular repairing and detecting on the equipment's is necessary. Regular maintenance is implemented by the professional and technical personnel, who manually examine each part of the equipment to see whether they operate normally. Assume that the manual inspection can find all the problems, then two specific maintenance rates occur. One is on-line maintenance rate, which occurs as the diagnosis programs detect and prompt the emergence of a failure, and the other is regular maintenance rate, which occurs during periodic detection and maintenance, and includes the testing time and repairing time. Compared with on-line maintenance rate, the regular maintenance rate is lower.

TABLE 1 Reliability calculation

R	R_{S1}	R_{S2}	R_{diff}
0.9000	0.88209000	0.98820900	0.10611900
0.9100	0.89531771	0.99057859	0.09526089
0.9200	0.90826168	0.99266093	0.08439925
0.9300	0.92090833	0.99446358	0.07355525
0.9400	0.93324418	0.99599465	0.06275047
0.9500	0.94525594	0.99726280	0.05200686
0.9600	0.95693046	0.99827722	0.04134676
0.9700	0.96825479	0.99904764	0.03079286
0.9800	0.97921616	0.99958432	0.02036817
0.9900	0.98980201	0.99989802	0.01009601
0.9950	0.99495025	0.99997475	0.00502450
0.9955	0.99545968	0.99997957	0.00451989
0.9960	0.99596813	0.99998387	0.00401574
0.9965	0.99647559	0.99998766	0.00351208
0.9970	0.99698205	0.99999095	0.00300889
0.9975	0.99748753	0.99999372	0.00250619
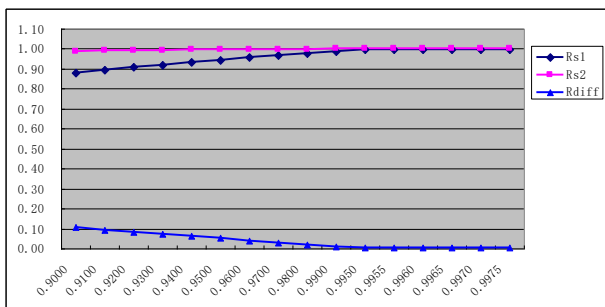
FIGURE 5 Reliability of fundamental and enhanced mode

The working principle of the enhanced mode is basically consistent with the fundamental mode, but it adds train-to-train direct channel to perform redundant communication channel to ensure the system safety. This kind of structure, as shown in Figure 6, adopts single machine performing double- channel software, and each channel possesses the diverse software, which makes the system realize 2-vote-2 from the hardware to the software. This ensures the system safety, further. This kind of the structure cannot only find most of the software errors in the process of operation, but also can detect the hardware failures. Therefore, its security is very high.
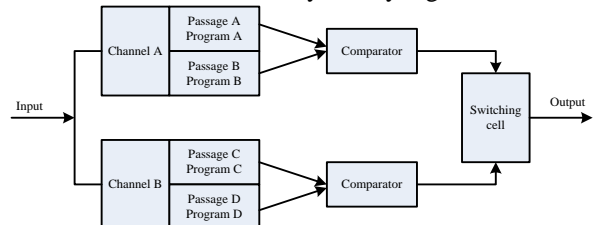
FIGURE 6 Reliability structure of the enhance mode

According to the former analysis, the Markov models of three kinds of dynamic redundant structures are isomorphic. Though observing carefully the states, we find two states meet the condition of mergers. Hence, we may establish a merged state transfer model for them as shown in Figure 7.

In Figure 7, the state zero expresses the two channels are perfect and work normally, and the state one expresses one channel is in failure and can be detected, and the state two represents one channel generates the failure (dangerous or safety) that self-diagnostic program cannot find but comparison program may detect out, and the state three presents the system safety failure, and the state four expresses the system dangerous failure but can be detected, and the state five presents the system dangerous failure undetected. In process of modeling, we assume that when one channel gets inspection and maintenance, and another channel also will have a chance to get detection and repairing. In addition, we also assume a maintenance rule of online repairing system under the condition that the system is not terminated. Likewise, we also assume that periodic detection and maintenance is perfect and can find any problem, and after repairing the system restores to the initial state. Thus, we may ignore the service arc from the state six to the state zero, as well as its service rate up. In fact, in terms of high security and high reliability system, it is meaningless to solve the steady state indexes.
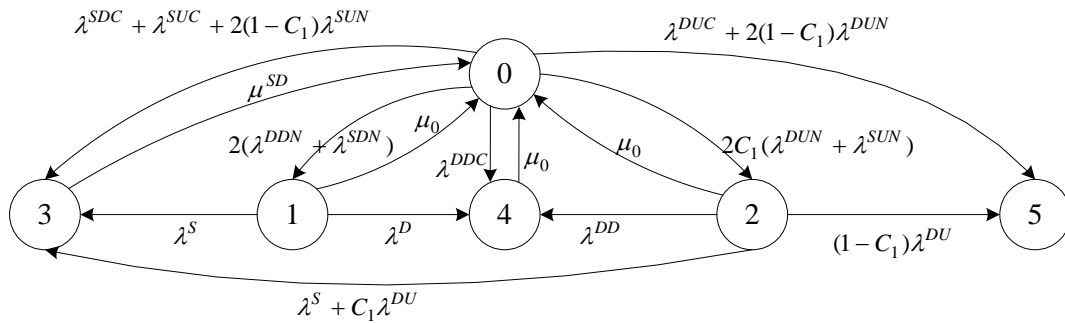
FIGURE 7 Merged isomorphic Markov model of enhance mode

According to Figure 7, we can acquire its state transition matrix P below

$$P = \begin{bmatrix} 1-\left(\lambda^{DC}+\lambda^{SC}+2\lambda^{DN}+2\lambda^{SN}\right) & 2\left(\lambda^{DDN}+\lambda^{SDN}\right) & 2C_1\left(\lambda^{DUN}+\lambda^{SUN}\right) & \lambda^{SDC}+\lambda^{SUC}+2(1-C_1)\lambda^{SUN} & \lambda^{DDC} & \lambda^{DUC}+2(1-C_1)\lambda^{DUN} \\ \mu_0 & 1-\left(\mu_0+\lambda^S+\lambda^D\right) & \lambda^S & \lambda^S & \lambda^D & 0 \\ \mu_0 & 0 & 1-\left(\mu_0+\lambda^S+\lambda^D\right) & \lambda^S+C_1\lambda^{DU} & \lambda^{DD} & (1-C_1)\lambda^{DU} \\ \mu_{SD} & 0 & 0 & 1-\mu_{SD} & 0 & 0 \\ \mu_0 & 0 & 0 & 0 & 1-\mu_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

In accordance with the matrix, the transient availability at any moment can be calculated based on Markov chain method in a detection cycle. Assume that the initial state is S_0, and the n-step state transient probability is p^n, and then the transient probability at n moment is $S_n = S_0 P^n$. According to S_n, we may solve PFD and PFS. The PFS is the probability of state three, and PFD is the probability sum of state four and five, and the availability is the probability sum in state zero, one, and two.

TABLE 2 Failure rates values

Failure rate type	Numerical value ($\times 10^{-5}$ h^{-1})
λ^{SDN}	0.98568
λ^{SDC}	0.07992
λ^{SUN}	0.10952
λ^{SUC}	0.00888
λ^{DDN}	0.24642
λ^{DDC}	0.01998
λ^{DUN}	0.02738
λ^{DUC}	0.00222

If the security failure rate is expressed by $\lambda^S = 1.184 \times 10^{-5}$ h^{-1}, the dangerous failure rate $\lambda^D = 0.296 \times 10^{-5}$ h^{-1}, and the diverse failure rates are listed in Table 2, 90% of the safety and dangerous failures can be found by self-diagnostic program, and the diagnostic coverage rate of comparison program is 99.9%, and online maintenance rate μ_0 equals 0.1. If the system generates a safety failure, then it could restart within 24 hours.

As $t = 8760$, under the conditions of the fundamental mode with $C_1 = 0$, the enhanced mode with $C_1 = 99.9\%$ we calculate the security and reliability indexes of the fundamental mode and the enhanced mode, as shown in Table 3.

TABLE 3 Security and reliability indexes values of diverse modes

Types	PFS($\times 10^{-5}$)	PFD($\times 10^{-4}$)	MTTF(h)
Fundamental mode	3.752355	6.814555	799033
Enhanced mode	3.351301	3.144774	894919

Seen from Table 3, the security and availability indexes of the enhanced mode are higher than the ones of the fundamental mode. For the hardware requirements, the enhanced mode is the highest, and from the hardware and software demands and technology implementation difficulty degree, generally, the fundamental mode is lower than the enhanced one. Hence, for modern railway train control system, it is very necessary to select the suitable redundant structures combining the requirements on safety and reliability, and cost, and the difficult-easy degree in technology realization.

5 Conclusion

In this paper, we have presented the supplement collision avoidance system for Chinese railway, namely Collision Avoidance System overlay Train Control System (CASOTCS), which is located completely on board and uses basically the information about position, speed and direction as well as additional data broadcasted by other CASOTCS with direct vehicle-to-vehicle radio communication. The on-board computer can analyze all the received context information, self-position and movement vector and an electronic track map to detect possible collisions, display an alert signal, and provide

the driver the most convenient strategy to avoid the potential danger. After the comparison of the reliability and safety between the traditional train control system and the enhanced CASOTCS, the reliability block diagram and isomorphic Markov model have been established. According to the relevant analysis and calculation, we have found that the enhanced train control system by adding collision avoidance system has the better reliability and safety than the traditional one.

In recent years, the Chinese railway systems have gone through a massive phase of upgrading and expansion, especially after the establishment of Chinese Train Control System (CTCS). Collision avoidance system has the ability to increase safety and efficiency.

The key issues to complete this system discussed in this paper will be further researched and a prototype should be realized to show the feasibility of the strategy.

Acknowledgments

References

[1] Carr Christopher 2014 *Railway safety performance in the European Union* http://www.era.europa.eu/ Document-Register/Documents/SPR2014.pdf/20 Sep 2014

[2] Strang T, Meyer zu Hörste M, Gu X 2006 Railway collision avoidance system exploiting Ad-hoc inter-vehicle communications and Galileo *World Congress and Exhibition on Intelligent Transportation Systems and Services (London, UK, 8–12 October)* **13** 1-8

[3] FAA 2011 *Introductions to TCAS II version 7.1* http://www.faa.gov/documentLibrary/media/Advisory_Circular/TCAS%20II%20V7.1%20Intro%20booklet.pdf/ 22 Sep 2014

[4] FAA 2011 *Automatic Dependent Surveillance-Broadcast (ADS-B)* http://www.faa.gov/nextgen/implementation/programs/adsb/22Sep 2014

[5] Navigation Center 2014 Automaic Systemtem Identification overview http://www.navcen.uscg.gov/?pageName=AISmain/ 22 Sep 2014

[6] Technical Committee 2007 *C2C-CC manifesto* http://www.car-to-car.org/ 22 Sep 2014

[7] He. T 2011 Safety analysis and design for the switch control unit of all-electronic computer interlocking system *TELKOMNIKA Indonesian Journal of Electrical Engineering* **10**(5) 1057-61

[8] Zhang. P, Zhao. Y 2003 Analysis on the reliability and safety of the interlocking control system of railway computer *China Safety Science Journal* **10**(4) 48-50

[9] Dong H, Ning B, Cai B, et al 2010 Automatic train control system development and simulation for high-speed railways *IEEE circuits and systems magazine* **10**(2) 6-18

[10] Ning B, Tang T, Qiu K, et al 2004 CTCS-Chinese Train Control System *WIT Transactions on The Built Environment* **74**(15) 1-7

[11] Garcia C R, Lehner A, Strang T, et al 2007 Comparison of collision avoidance systems and applicability to rail rransport *Proceedings of the 7th International Conference on Intelligent Transportation System Telecommunication* **7** 521-6

[12] Lehner A, Strang T, Garcia C R 2008 A reliable surveillance strategy for an autonomous rail collision avoidance system *Proceedings of the 15th ITS World Congress* **15** 81-8

[13] Strang T, Lehner A, Garcia C R 2009 Building a 2nd line of defense for safer rail *EU RAIL mag Business & Technology* **19** 92-5

[14] García C R, Strang T, Lehner A A broadcast vehicle to vehicle communication system in railway environments *Proceedings of the 1st Annual International Symposium on Vehicular Computing Systems* **1** 70-6

[15] Kim M C 2011 Reliability graph with general gates: an intuitive and practical method for system reliability analysis *Annals of Nuclear Energy* **38**(11) 2456-61

[16] Goble W M 2010 *Control System Safety Evaluation and Reliability 3rd Edition* RaIeigh: ISA

[17] Su H S 2013 Reliability and security analysis on two-channel dynamic redundant system *TELKOMNIKA Indonesian Journal of Electrical Engineering* **11**(5) 2594-604

[18] Ebeling C E 2008 *An Introduction to Reliability and Maintainability Engineering* Tsinghua Unversity Press, Beijing

[19] Lin J T, Dang J W Reliability Models of GSM-R Redundant Network on High-Speed Railway *Applied Mechanics and Materials* **198** 1783-8

Author

Junting Lin, born in August, 1978, Lanzhou City, Gansu Province, P.R. China

Current position, grades: the Lecturer of School of Automation & Electrical Engineering, Lanzhou Jiaotong University, China.
University studies: received his B.Sc. in Automatic Control from Lanzhou Jiaotong University in China. He received his M.Sc. in Communication and Information System from Beijing Jiaotong University in China.
Scientific interest: His research interest fields include Train Control System, Vehicle to Vehicle communication
Publications: more than 50 papers published in various journals.
Experience: He has teaching experience of 12 years, has completed three scientific research projects.

Xiaoming Wang, born in November, 1954, Lanzhou City, Gansu Province, P.R. China

Current position, grades: the Professor of School of Automation & Electrical Engineering, Lanzhou Jiaotong University, China.
University studies: received his B.Sc. in Railway Signal Control from Lanzhou Railway Institute in China. He visited Ryerson Institute of Technology as senior visiting scholar.
Scientific interest: His research interest fields include Railway Signal Control System, Computer-based Interlocking system.
Publications: more than 160 papers published in various journals.
Experience: He has teaching experience of 52 years, has completed twenty six scientific research projects.

Jianwu Dang, born in May, 1963, Lanzhou City, Gansu Province, P.R. China

Current position, grades: the Professor of School of Automation & Electrical Engineering, Lanzhou Jiaotong University, China.
University studies: received his B.Sc. in Railway Signal Control from Lanzhou Railway Institute in China. He received his M.Sc. in Computer Science from Southwest Jiaotong University in China. He received his Ph.D. in Communication and Computer Engineering from Southwest Jiaotong University in China.
Scientific interest: His research interest fields include Train Control System, Intelligent Transportation System and Fuzzy Neural Network.
Publications: more than 170 papers published in various journals.
Experience: He has teaching experience of 28 years, has completed fifty scientific research projects.

Short-Term photovoltaic system power forecasting based on ECSVM optimized by GA

Yun-jun Yu*, Yun-tao Xue, Sui Peng, Chao Tong, Zi-heng Xu

College of Information Engineering, Nanchang University Xuefu Rd. 999, Nanchang, Jiangxi, 330031, China

Abstract

It is of great significance to research PV forecasting techniques for mitigating the effects of the randomness of the Photovoltaic output. This paper analyses many factors from PV which impact photovoltaic output and extracts the main factors, forming sample data combined with the historical database generation data from PV monitoring system. And an error correction SVM method (ECSVM) is used to calculate the open integration of photovoltaic power storage system in advance or after the time in order to try to eliminate the system error between the predicted and actual values. At the same time, using genetic algorithm to optimize kernel function parameter and the error penalty factor and other parameters in this model, the establishment of the GA-ECSVM model improves portfolio optimization model parameter prediction accuracy and efficiency of the selected type. Finally examples verified and compared with standard SVM methods and ECSVM method, predicting effects show that: The GA-ECSVM optimization model presented in this paper has better learning ability and generalization ability in the short term prediction of photovoltaic power generation, with the prediction accuracy of 95.2016%.

Keywords: PV, SVM, error correction, GA, forecast

1 Introduction

Along with the intensity of environmental pollution caused by global energy shortage and the use of fossil resources, solar power as one of the important renewable energy, has received wide attention [1, 2]. Photovoltaic system has more practice value in this area because of its technical improvement. Because of the stochastic weather and fluctuation of illumination intensity, it will inevitably influence generation scheduling greatly [3-5]. The accurate PV power generation forecast is one of the effective means to improve the PV power capacity, the stability and economy of the power grids operation.

The theory of PV power generation forecasting has been researched in recent years, and a lot of forecasting methods are proposed. In [6], using historical weather data sources to make predictions about illumination intensity in future. But the accuracy of PV power generation prediction needs to be improved to satisfy the actual production. In [7], a back propagation (BP) neural network forecasting model was proposed whose input parameters were ambient temperature, humility and cloud cover, which improve accuracy of PV power generation prediction. But photovoltaic (PV) generation forecasting models need to take cloud cover as their input parameters. However, they were difficult to implement in China due to insufficient weather stations available.

Extraction temperature, Solar Radiation Intensity and sunshine time were the main factors, which affect PV power generation. This paper provides error correction SVM method (ECSVM) according to historical data. But error correction SVM method (ECSVM) is still deficits in the technical parameter selection. In allusion to these problems, this paper use genetic algorithm to optimize kernel function parameters and the error penalty factor and other parameters in this model, the establishment of the GA-ECSVM model improves portfolio optimization model parameter prediction accuracy and efficiency of the selected type.

2 Important factors affecting PV generation

For different districts and locations, meteorological factors, including solar irradiance, temperature, and so forth, are always changing [8, 9]. In order to efficiently utilize renewable energy using solar energy, an analysis of the characteristics of meteorological factors at a potential site should be considered. Data from the solar energy grid-connected PV plants, including the amount of solar electricity being produced and the meteorological factors, will be made available to analyse.

The output power of photovoltaic power station on sunny days is presented in Figure 1. It has strong randomicity, but similar on the whole.
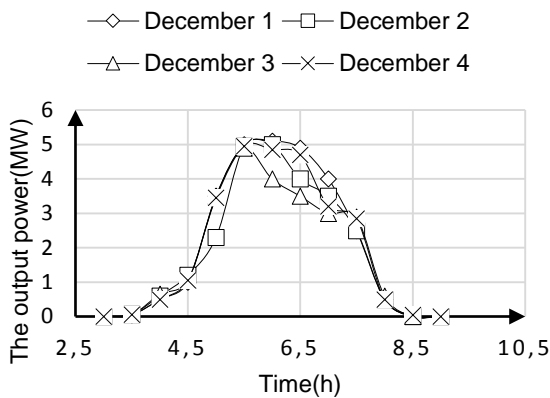
FIGURE 1 Daily power output in sunny days

FIGURE 2 Solar radiation and power output of PV system

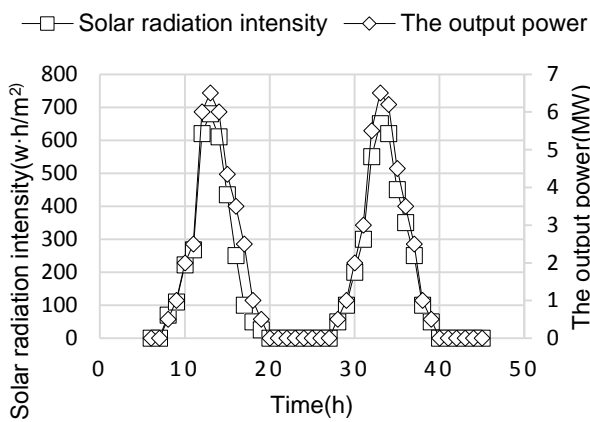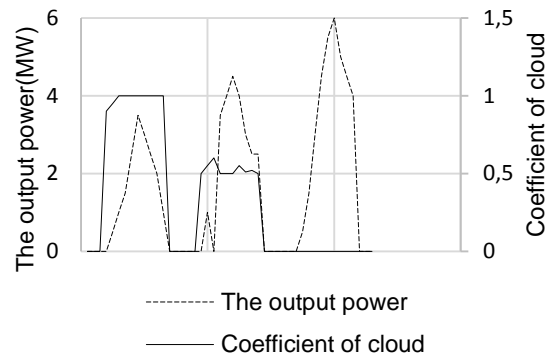FIGURE 3 Coefficient of cloud and power output of PV system
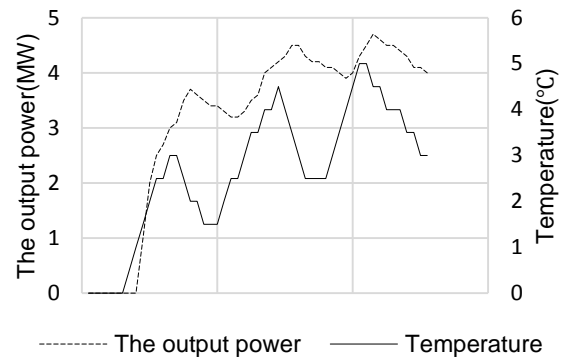
FIGURE 4 Temperature and power output of PV system

2.1 SOLAR RADIATION INTENSITY

As shown in Figure 2, The intensity of solar radiation has a significant impact on photovoltaic power generation. In photovoltaic system, the output of photovoltaic battery is affected greatly by the Solar Radiation Intensity, it has obvious non-linear features [10]. Thus, the intensity of solar radiation will be used as one significant impact of the predictive model.

2.2 CLOUD

The effects of the cloud on solar radiation is considered. The response time of a PV (photovoltaic) plant is very short and its output power follows the abrupt change in solar irradiance level due to alternate shadow by clouds [11, 12]. As shown in Figure 3, When the cloud covers a large area, the output power will be reduced relatively, also the shadow by clouds will be used as one significant impact of the predictive model.

2.3 TEMPERATURE

The change rules of the output power were similar with the change tendency of temperature [13, 14]. As shown in Figure 4, there is a positive correlation between temperature and the output power. In addition, the temperature will be used as one significant impact of the predictive model.

3 The theory of SVM

The advantage of support vector machine predictive model can overcome the shortcoming of traditional methods, which are over-fitting, non-linear, disaster of dimensionality, local minimum.

3.1 THE THEORY OF SVM REGRESSION

Regression problems was based on a new input stylebook data x to inference corresponding output y.

Data sample set is $\{(x_i, y_i),...,(x_l, y_l)\}$, where $x_i \in R^n, y_i \in R, i = 1, 2, 3,..., l$.

The basic idea of SVM estimate regression means, performing the data of input space into a high-dimensional feature space through non-linear mapping relationship.

The value of sample data $\{x_i, y_i\}$, $i = 1, 2, 3,..., s(x_i \in R^n, y_i \in R)$, where y_i is the exception, s is the total number of data points.

Regression problem is solved by Loss function in SVM. Use function: $y = f(x) = [w\varphi(x)] + b$.

Take the extreme value of optimization goal:

$$\min Q = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \zeta_i), \tag{1}$$

$$s.t.\begin{cases} y_i - [wk\varphi(x_i)] - b \leq \varepsilon + \xi_i \\ [wk\varphi(x_i)] + b - y = \varepsilon_i + \zeta_i \\ \xi_i, \zeta_i \geq 0, i = 1, l, s \end{cases},$$

where C is error penalty factors; ξ_i and ζ_i are relaxation factor; ε is loss function. Loss function can show decision function through sparse data points. The loss Equation (2) introduced has a good effect.

$$L_\varepsilon(y) = \begin{cases} 0 & |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & |f(x) - y| \geq \varepsilon \end{cases}. \tag{2}$$

Using Lagrangian multiplication a_i and b_i, which facilitate convex optimization problems into quadratic maximum.

TABLE 1 Common kernel function

	Name	Expression
1	Liner kernel function	$k(u,v) = (u \cdot v)$
2	Polynomial kernel function	$k(u,v) = (r(u \cdot v) + coef\,0)^d$
3	RBF kernel function	$k(u,v) = \exp(-r|u-v|^2)$
4	Sigmoid kernel function	$k(u,v) = \tanh(r(u-v) + coef\,0)$

Note: the r is kernel function parameter.

Support vector machines(SVM) method is developed for solving highly nonlinear classification, whose program are mapped into a high-dimensional feature space through a certain transformation function to nonlinear data variables and then from line regression of high dimension using kernel function $k(x, x_i) = \phi(x)\phi(x_i)$. Introduce kernel function to replace inner product computation. optimization goal integrated as shown below :

$$\max W(a_i, b_i) = \sum_{i=1}^{n} y_i(a_i - b_i) - \varepsilon\sum_{i=1}^{n}(a_i + a_i') -$$
$$\frac{1}{2}\sum_{i,j=1}^{n}(a_i - a_i')(a_j - a_j')(x_i - x_j'), \tag{3}$$

$$s.t.\begin{cases} \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i' \\ 0 \leq a_i, a_i' \leq C \end{cases} \quad i = 1, 2, ..., n,$$

where C is used to control the complexity of the model and compromises approximation errors. When the C larger, the fitting degree higher; ε was used to control regression approximate error and generalization ability models.

3.2 THE METHOD OF PREDICTING

The time series are $\{x_1, x_2, x_3, ..., x_n\}, i = 1, 2, 3, ...n$ and predictive goal value is $\{x_n\}$. This paper set up the functional projective relationship which is $R^m \to R$ from the input to the output .m is embedded dimension. We can get samples for learning SVM:

$$X = \begin{bmatrix} x_1 & x_2 & L & x_m \\ x_2 & x_3 & L & x_{m+1} \\ L & L & L & L \\ x_{n+m} & x_{n+m+1} & L & x_{n-1} \end{bmatrix}, Y = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ M \\ x_n \end{bmatrix}.$$

Regression function:

$$y_i = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_j}) + b, j = m+1, ..., n. \tag{4}$$

The first step:

$$x_{n+1} = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_{n-m+1}}) + b, \tag{5}$$

where $\overline{a_{n-m+1}} = \{x_{n-m+1}, x_{n-m+2}, ..., x_n\}$.

Then:

$$\overline{a_{n-m+2}} = \{x_{n-m+2}, x_{n-m+3}, ..., x_n, x_{n+1}\}.$$

Therefore:

$$x_{n+2} = \sum_{i=1}^{n-m}(a_i - a_i')k(x_i, x_{n-m+2}) + b. $$

Result:

$$x_{n+l} = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_{n-m+l}}) + b, \tag{6}$$

where $x_{n-m+l} = \{x_{n-m+l}, x_{n-m+l+1}, ..., x_{n+l-1}\}$.

3.3 PV GENERATION PREDICTION MODEL BASED ON ECSVM

PV system affected by the environment, there are many uncertain disturbance sources. The errors between predictive value and the actual value were unavoidable. In order to minimize error, error integral method is proposed. Its characteristics are as follows:

$$Y_y = [x(1)x(2)x(3)\ldots x(120)], \tag{7}$$

$$Y_s = [y(1)y(2)y(3)\ldots y(120)], (n \leq 120). \tag{8}$$

Y_y – predictive power, Y_s – real generation power,

$$E_r = \sum_{i=t}^{n}[Y_y(i) - Y_s(i)], (t \leq n \leq 120), \tag{9}$$

where $Y_y(i)$ – real generation power at time i equals; – Predictive power at time $\underline{i}$ equals; E_r – the errors between predictive value and the actual value.

$$t = \frac{E_r - E_g \times \eta_{inv}}{P_d \times \eta_r} , \qquad (10)$$

where t – energy storage opening time error; E_g – the errors between predictive value and the actual value; P_d – energy storage equipment output power; η_{inv} – efficiency of PV power generation system; η_r – efficiency of energy storage output energy.

4 PV Generation prediction model based on GA-ECSVM

SVM can have different performances of classification through choosing different Kernel Functions and parameters [15, 16]. The performance of the SVM is influenced by kernel function parameter r [17, 18]. Meanwhile, kernel function is the key technology of SVM. Using different kernel function will affect the learning ability and generalization ability of SVM. In view of the present insufficiency in the selection technology aspect of SVM, several proposals were put forward, which will improve SVM learning ability, generalization performance and the ability to choosing the kernel function and function's parameters.

Genetic Algorithm is based on the nature selection and genetic transmission mechanisms, whose advantages is high collateral, stochastic, self-reliance [19, 20]. To improve the SVM, ECSVM optimized by GA are given in Figure 5.
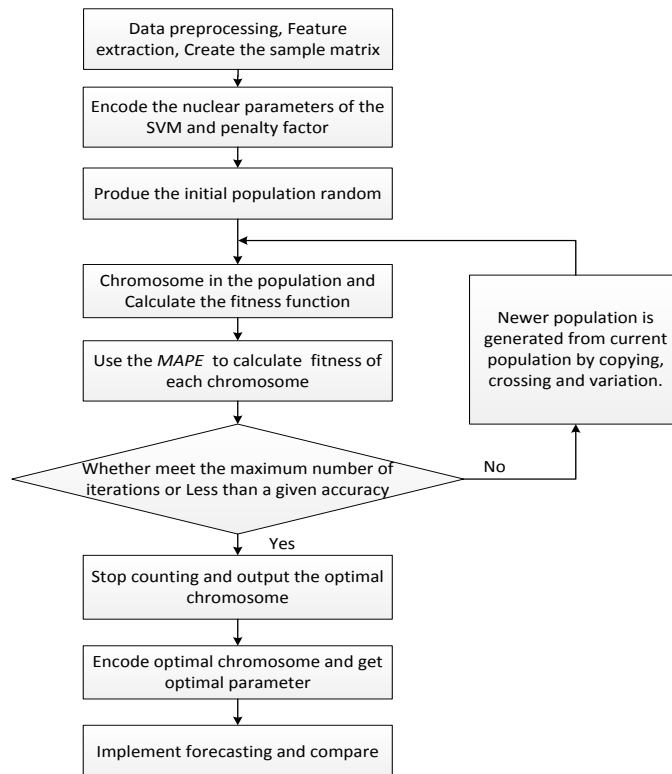
FIGURE 5 Flow chart

FIGURE 6 Decoding process

Step 1: In PV power generation data processing, the programme checks and modifies the bad-data.
Step 2: The factors which affect PV power generation multiple. The main factors, such as solar radiation intensity, sunshine time, and temperature, are considered in the paper.
Step 3: Consider the couple role of above factor. We can get an input matrix.

Step 4: The training set data and test data consist of kernel function and penalty factor, which is produced by a population of randomly generated each individual genes in the string of decoding, were input into model for training and simulation. Decoding process as show in fig. (6).

Step 5: The mean absolute percentage error of test sample *MAPE* is considered mean function to evaluate the population:

$$MAPE = \frac{1}{N} \sum_{k=1}^{N} \frac{\left| p_{simu.k} - p_{trag.k} \right|}{p_{trag.k}} \times 100\% , \qquad (12)$$

Step 6: The next generation of individuals were choose based on the individual fitness and selection principle.

Step 7: Selection, crossover and mutation operations were used to produce the next generation.

Step 8: judging whether the terminal condition is meet, if meet, turn into the next step, otherwise go to step 5.

Step 9: The best individual were exported and considered as its proximate optimum solution in this problems.

Step 10: Then, the forecast is realized by the use of Genetic algorithm that used to derive the corresponding kernel function , kernel function parameter and error warning factor.

Step 11: MAPE is used to predict and evaluate the performances, and last, they were compared with the traditional methods.

5 Case study

PV monitoring system historical electricity generation data and environmental parameters are included in the sample data of prediction system. Experimental samples are selected from Nanchang University rooftop photovoltaic power generation data, which recorded for 5 min interval. Foe the training sample the period from August 1 2011 to August 6 2011 was selected. The data from August 7 2011 are tested as forecast samples.

A quick look at the Table 2 above indicates that forecast accuracy and training accuracy of GA-ECSVM is better than ECSVM. Forecast effect of GA-ECSVM is given in Figure 7, which has high accuracy of prediction. In order to verify the superiority of the GA-ECSVM, Comparative forecast effect of SVM, ECSVM and GA_ECSVM can help rank forecast accuracy. The error of measurement using different methods are compared in Figure 8.

From Figure 8 the accuracy of prediction of GA-ECSVM model is higher than SVM and ECSVM. In order to calculate the errors between predictive value and the actual value accurately, mean absolute error method are proposed.

From Table 3 we know that ECSVM model, which has optimized the GA is more exactitude. The relative prediction error precisions are all below 5%.

TABLE 2 Contrast GA-ECSVM with ECSVM predictions

Experiment	SVM parameters			Training accuracy (%)	Forecast accuracy (%)
	Kernel function type	Kernel function parameter *r*	Penalty factor		
Experience choice	1		1	62.5	86.67
	1		10	72.5	83.33
	2	1	10	95	90
	2	10	1	86	56.67
	3	1	10	95	86.67
	3	10	1	94	73
	4	1	10	57.5	73.33
	4	10	1	52.5	33.33
GA optimization	2	7.375	11.91	98	96
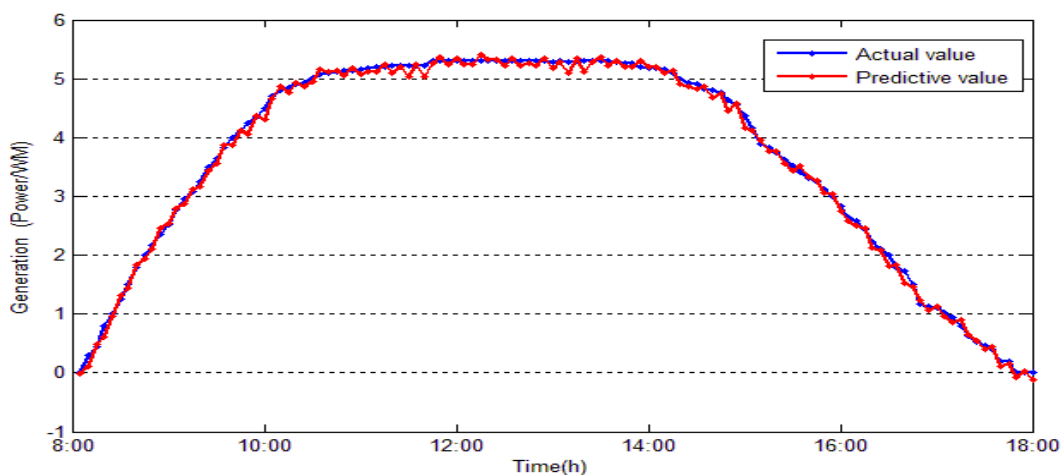
FIGURE 7 The power generation forecasting
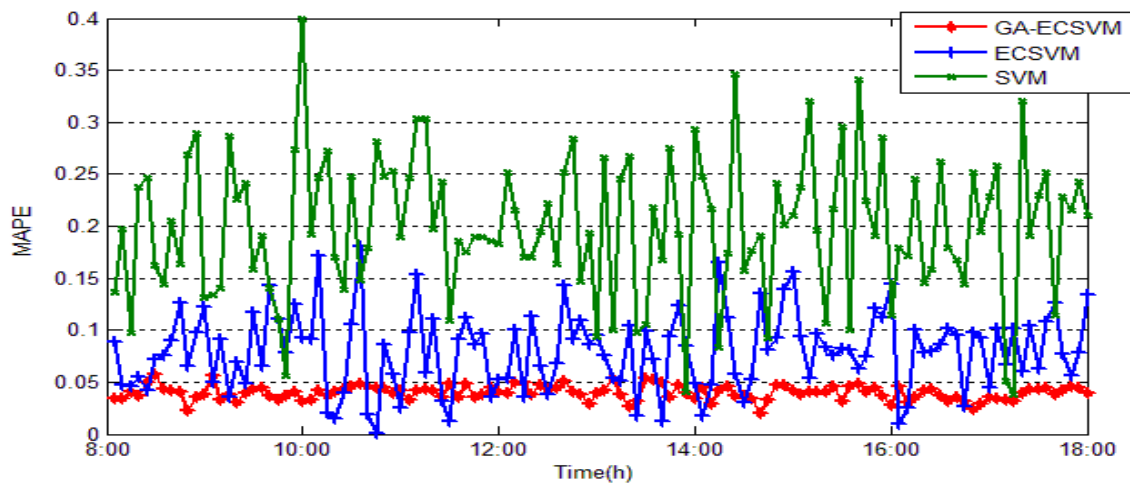
FIGURE 8 Forecasting error curve

TABLE 3 Forecasting error

	GA-ECSVM	ECSVM	SVM
MAPE (%)	4.7984	9.6255	15.3654

6 Conclusions

Power forecasting has received a great deal of attention due to its importance for planning the operations. In this paper, the GA-ECSVM model is proposed to forecast the output of PV power generation. Firstly, an error correction SVM method (ECSVM) is used to calculate the open integration of photovoltaic power storage system in advance or after the time in order to try to eliminate the system error between the predicted and actual values. Meanwhile, using GA to optimize kernel function parameters and the error penalty factor and other parameters in this model, GA-ECSVM model improves portfolio optimization model parameter prediction accuracy. Finally, an application example shows that : the GA-ECSVM model has better learning ability and generalization ability in the short term prediction of photovoltaic power generation than SVM and ECSVM, with the prediction accuracy of 95.2016%.

Acknowledgements

References

[1] Liu Q 2000 Strategic Choice for National Adjustment of Energy Structure *Advance In Earth Sciences* **15**(2) 154-64
[2] Xue F, Chen K, Wang N 2011 Coordinated Control Frame of Large-scale Intermittent Power Plant Cluster *Automation of Electric Power System* **35**(22) 45-33 *(in Chinese)*
[3] Zhang L, Hurley W G, Wöfley W H 2011 A new approach to achieve maximum point tracking for PV system with a variable inductor *IEEE Transactions on Power Electronics (Special Issue on Microgrids)* **26**(4) 1031-7
[4] Li J, Xu H, Zhao H 2008 Dynamic modeling and simulation of the grid-connected PV power station *Automation of Electric Power Systems* **32**(24) 83-7 *(in Chinese)*
[5] Chakraborty S, Weiss M D, Simoes M G 2007 *IEEE Transactions on Industrial Electronics* **54**(1) 97-109
[6] Yona A, Senjyu T, Funabashi T 2007 *IEEE Power Engineering Society General Meeting* 2007 1-6
[7] Dai Q, Duan S, Cai T, Chen C, Chen Z, Qiu C 2011 Short-term PV generation system forecasting model without irradiation based on weather type clustering Proceedings of the CSEE **31**(34) 28-35 *(in Chinese)*
[8] Lu Jing, Zhai H, Liu C, Wang X 2010 Study on Statistical Method for Predicting Photovoltaic Generation Power *East China Electric Power* **38**(4) 563-7 *(in Chinese)*
[9] Yang D, Wang Y, Jiao Y 2013 Generation forecasting for photovoltaic system based on wavelet neural networks *Renewable Energy Resources* **31**(7) 1-5 *(in Chinese)*

[10] Sharan A M 2009 Efficiency enhancement of stationary solar energy based power conversion systems in Canada *Applied Energy* **86**(9) 1405-9
[11] Dai Q, Duan S, Cai T, Chen C, Chen Z, Qiu C 2011 Short-term PV generation system forecasting model without irradiation based on weather type clustering Proceedings of the CSEE **31**(34) 28-35 *(in Chinese)*
[12] Li Y, Wang R, Wang T, Wu J, Xu Y 2007 Experimental Performance Analysis and Optimization of A Direct Expansion Solar Assisted Heat Pump Water Heater *Acta Energiae Solaris Sinica* 2007 **28**(5) 464-71 *(in Chinese)*
[13] Chen C, Duan S, Cai T 2011 Short-Term Photovoltaic Generation Forecasting System Based on Fuzzy Recognition Transactions of China Electrotechnical Society **26**(7) 83-9 *(in Chinese)*
[14] Armstrong A, Hurley W G 2010 A thermal model for photovoltaic panels under varying atmospheric conditions *Applied Thermal Engineering* **30**(12) 1488-95
[15] Mao M, Gong W, Chang L 2013 Short-term photovoltaic forecasting based on EEMD-SVM combined method. Proceeding of the Chinese Society of Electrical Engineering 2013 **33**(34) 17-24
[16] Fukushi D, Shichiri M, Sugiyama S, Yoshino T, Hagiwara S, Ohtani T 2013 Scanning near-field optical/ atomic force microscopy detection of fluorescence in situ hybridization signals beyond the optical limit. *Exp. Cell Res* **289** 237-44

[17] Chapelle O, Vapnik V, Bousquet O, Mukherjee S 2002 Choosing Multiple Parameters for Support Vector Machines *Machine Learning* **46**(1) 131-59

[18] Huang C, Lee Y, Lin D, Huang S 2007 Model selection for support vector machine via uniform design *Computational Statistics & Data Analysis* **52**(1) 335-46

[19] Saini L M, Aggarwal S K, Kumar A 2010 Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market *IET Generation, Transmission & Distribution* **4**(1) 39-49

[20] Cheng G, Wang H, Sun X 2012 Model selection for SVM classification based on kernel prototype and adaptive genetic algorithm *Journal of Graduate University of Chinese Academy of Sciences* **29**(1) 62-9

Authors

Yunjun Yu, born in June, 1978, Jiangxi, China

Current position grades: Doctor, lecturer at Nanchang University, Nanchang, Jiangxi, China.
University studies: Automation in Nanchang University. Ph.D degree at Chinese Academic of Science, 2013.
Scientific interest: photovoltaic inverter, fault diagnosis, micro-grid.
Publications: 30 Papers.

Yuntao Xue, born in December, 1991, Henan, China

Current position, grades: Junior student at Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China
Scientific interest: power system and its automation.

Sui Peng, born in November, 1992, Jiangxi, China

Current position, grades: Junior student in Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China.
Scientific interest: power system and its automation

Tong Chao, born in July, 1992, Jiangxi, China

Current position, grades: Junior student in Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China
Scientific interest: power system and its automation

Ziheng Xu, born in September, 1992, Nanchang, China

Current position, grades: Junior student in Nanchang University, Jiangxi, China.
University studies: Computer science and technology in School of Electronics and Information Engineering.
Scientific interest: optimization and application of the algorithm.

Research on the application value of computer technology in hotel management

Wensu Xu*, Leilei Wang

Qingdao Hotel Management College, Qingdao, Shandong, 266100, China

Received 12 June 2014, www.cmnt.lv

Abstract

Hotel management become much important with the increase of services of hotel. And the traditional hotel management pattern have not adapted to the modern hotel management. Therefore, management with high-tech technology is needed in modern hotel. Computer application can well solve management mode of modern hotel and improve the service level of hotel. With the development of IT technology and hotel industry, current hotel management have developed into a new age and new systematic development direction emerges constantly. This paper mainly discusses the application value of computer technology in hotel management. And it illustrates the importance of computer technology in hotel management by explaining the systematic design and implement of hotel management.

Keywords: hotel management, computer, information system

1 Introduction

Anthropogenic flow scale expands with social progress. Large anthropogenic flow amount promotes the rapid development of hotel industry and improves the requirement of service quality. Thus, management pattern should keep pace with age progress and high technology is needed in hotel management to better solve the value and economic benefit of hotel. Application of computer technology in hotel management solves the problem. Based on information technology and computer technology, hotel management system achieves the management of manpower, physics, capital, logistics and information.

Application of computer information technology will not decrease service to customer. It can improve working efficiency in hotel service operation level and make stuffs have more time to provide more service for customers. It can directly improve service quality and provides more prefect and comfortable experiment for customer in service product. It can provide better tools and improve marketing effect in hotel marketing area. It greatly enhances the management control and marketing prediction and analysis ability by establishment of rapid communication channel [1], design work process and automatic monitoring and feedback analysis of database flow.

More and more hotels tend to combine emerging network technology with the rapid development of network technology. A large amount of manpower and material resources are put in to develop internet booking business. E-commerce war begins secretly. Portal system of hotel industry is facing with the huge challenge.

Therefore, hotel managers must realize that current hotel portal system should not only fully advertise their brand in network space, but also meet different requirement of customers. Different customers have different purchasing behaviour. Hotel portal system should meet different requirement of consumption as possible. Hotel need provide more personalized and diversity service to improve customer satisfaction for transforming potential customers into practical customers and cultivating loyal customers. For instance, Hilton Hotel Corporation has near one hundred years of history. It owns 520 subbranches and 475 thousands of guest room in the world. It can also establish portal systems with regional characteristics adopting different management principle according to the cultural background of different area [2]. In addition, Yinfeng Days Inn Business Hotel is a business hotel managed by American Days Inn Hotel Corporation. Yinfeng Days Inn is characterized by far geographic position, quiet and comfortable environment and low price. The system can be clear at a glance and shorten the time of browsing system.

2 Overview of Relative Technology

2.1 COMPUTER TECHNOLOGY

The involved computer technology is mainly computer networking technology. Computer network is the product combining computer and communication technology. It can connect the scattered computer users together by transmission media. With the support of relative communication software, computer users can share

* *Corresponding author* e-mail:15697479@qq.com

computer equipment, application software and data and realize the transmission of voice band and video.

Network of current hotel management information system can share server, for example, Novell NetWare and NT72000 server. Server is a computer. Its operation system can make multiple users to visit the software. Meanwhile, it can provide services such as network printer management and fax service management. In this network, a computer accepts servers are termed as client. Client can communicate with server and acquire various service application provided by server. Information transfer is shown in Figure 1.
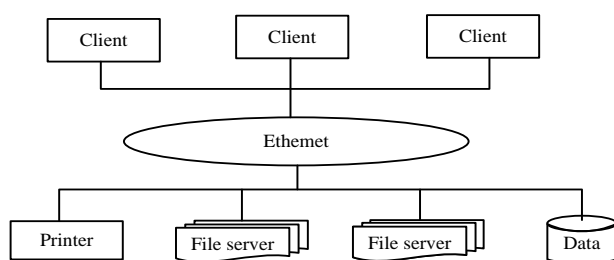
FIGURE 1 Schematic diagram of computer network structure

For guest room booking, system mainly adopts radio interface technology, which is composed of series transmitting entity between exchanger interface and network interface of relative user. It provides wireless implementation systems that have transmission capacity for transmitting telecommunication services. Computer technology is also needed in PSTN access technology and digital data network access technology.

2.2 DATABASE TECHNOLOGY

Core technology applied in systematic development is ASP+SQL Server 2000. It is a hotel management information system based on B/S pattern [3]. Design of database should according to the practical demand of hotel management. The following designs are mainly fulfilled:

(1) Design of user's jurisdiction. Distributed database model is established for the overall planning of hotel management information system. Users can share the information in database record.

(2) Design of index file. The information submitted by clients should be quick positioned and accurately found for managing a large amount of data information of client network booking for guest room.

(3) Design of file and format. In practical index file, database table should be established for administrators to accurately find client information.

(4) View design. View is established based on the name of client, room number and ID number. Inquire of this information or other operation can be carried out on view.

2.3 SOFTWARE FRAME TECHNIQUE

The applied B/S architecture mainly applies WWW browser technology that is constantly mature combining with multi-scripts language. It realizes strong function, which can only achieved by complicated special software in the past and saves development cost [4]. System installation, modification and maintenance all rely on server, which brings convenience for hotel management and also saves maintenance cost of systems technology. Administrator need only one browser to operate all modules. It really achieved function of zero client side and is easy to update automatically in runtime. Meanwhile, software frame technology of B/S also provides heterogeneous machine, heterogeneous network and online of heterogeneous application service. It combines with computer radio interface technology, which is convenient for different clients to fulfil guest room booking in different places to a larger extend.

3 Design of Hotel Management Information System

3.1 DEMAND ANALYSIS OF HOTEL MANAGEMENT INFORMATION SYSTEM

In current hotel management information system, system content is easy to operate and structure is simple. It mainly has function bar, booking system and propaganda picture [5]. Browser is easy to find the information and operation is simple and effective. Client only needs to input user name, password and real name when booking guest room online and then fulfil booking. Client can check whether the booking is completed any time. The booking is usually submitted in form. The whole processes only need one conversion process. Then the basic information of user can be stored and it is convenient for administrator to management client's information. Hotel administrator only need to input username and password for login and then manage the relative information of client. There can be more than one administrator. Administrators can be added in admin table.

Current hotel management information system does not have enough multi-media factors [6]. A good hotel management information system needs to insert video and voice band in system page. Streaming media technology adopts buffer storage to realize basic technology of streaming transmission. It can not only show real time dynamic information of hotel to clients but also integrate specialty catering and tourism of the located area of hotel to meet clients' individual demand. Database safety should be improved to prevent the reveal or damage of clients' information. The basic operation principle of streaming media technology is shown in Figure 2.
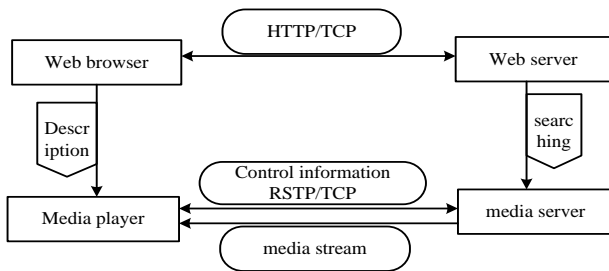
FIGURE 2 Operation principle of streaming media transmission

Client information needs to be encrypted for safety. Encryption technology mainly adopts DES algorithm [7]. In addition, network access control, network jurisdiction space, network server safety control, safety control of network node and port are also need to be carried out. Firstly, access of database need to be controlled for the safety and integrity of database. Then recovery technology is applied. And DBA store data into another disk at fixed period. When the database is damaged, it can help to recover database operation. Working operation flow diagram of hotel management information system is shown in Figure 3.
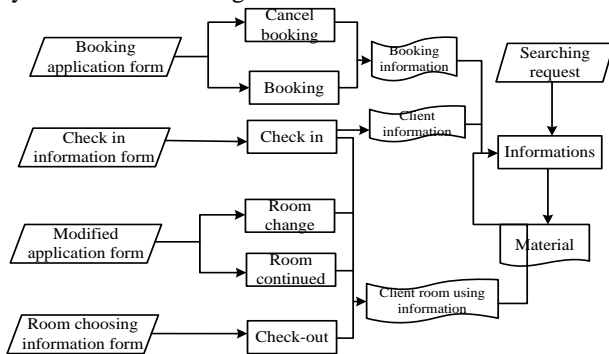
FIGURE 3 Flow diagram of hotel management information system

3.2 OVERALL ARCHITECTURE OF HOTEL MANAGEMENT INFORMATION SYSTEM

The architecture of system software mainly has three layers: data layer, web service and dedicated browser, as shown in fig.4. Data layer, that is, database, is data that stored in certain form in computer storage device. In hotel management information system, clients' information should be stored for later further searching after storage and collection. It is convenient for system administrator to process and obtain needed information. With the rapid development of computer technology, people now have changed storing data in file cabinet. People store and manage large amount of complicated information scientifically with computer and database technology. These precious information resources can be conveniently applied and managed, which improves the efficiency of hotel management and lower management cost [8]. Data stored in the database is organized, described and stored according to certain data model. These data have small redundancy, high data independence and expansibility and can be shared by users.
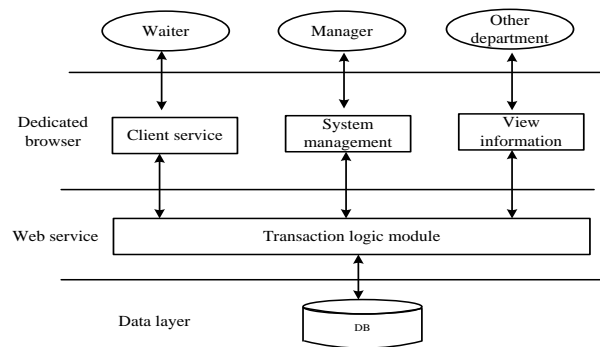
FIGURE 4 Architecture diagram of system software

Management and control system of guest room are composed of backend database, guest room intelligent controller, supervisory control computer in every floor, guest room server, power engineering server, reception desk supervisory control computer and other accessory equipment. System structure adopts second level distributed network structure and applies C/S star network topological structure between supervisory control computer of every floor and guest room server. It adopts TCP/IP protocol to communicate, links by fast Ethernet and achieve by Winsock. Lower computer is the second level system of controller. Upper computer is PC. RS232 full-duplex serial port is used for communication between lower computer and upper computer [9]. Guest controller is lower computer responsible for field control and information collection. Floor supervisory control computer is upper computer. It monitors all guest rooms in the floor in real time and achieves modification of demonstration and control parameter and process of information.

Hotel management information system contains various functions. Every function has several correlation subfunction modules.

(1) "Administrator management" function module is used for managing the account of administrator such as adding administrator's account, modifying password, delete account.

(2) "System management" function module is used for management system allocation, system notice and database backups. Allocation of management system is mainly used for controlling the demonstration of information in system. These information include demonstration pieces, number of character, limitation of file size, storage of file catalogue, catalogue of backend management file, permitted file type, delete operation supported by FSO and effective time of backend login. System notice is used for managing, modification and deletion of title and release of notice content. Backups of database mainly includes backup of system data file and data storage path.

(3) "Enterprise information management" function module is mainly used for managing hotel information such as hotel introduction, manager speech, hotel course, hotel honour, location, etc. It has function of viewing, modification and deletion. Meanwhile, new column such as content and character can be added in this module.
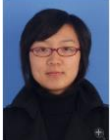
(4) "Service management" function module is mainly used for managing service item classification, service item management, service item addition, service item review, room price addition, room information management and room booking management.

(5) "News management" function module is mainly used for all news management and news classification management. All news management is mainly focus on modification or deletion of tourism news, industry information and enterprise news. News addition mainly includes addition of news title, news content and news classification.

(6) "Message management" function module is used for inquiring, replying or deleting message.

A basic hotel management system database contains multiple tables, which respectively store data information of relative subfunction. "Guest room information" and "guest information" are expressed as critical tables, which are used for storing the basic information of guest room and information of check in client. Other table involving guest room are just used for recording relative guest room number: "guest room information".

Design of backend management system includes administrator information, system information, service management, news management and message administrator. Administrator can enter the backend management module of hotel management information system by input user name, password and verification code. Administrators mainly have two operations in backend module: one is the basic setting of room information; second is the modification and deletion of room historical material module.

Function of database backup is backup of data file of enterprise system management. Current database path can be showed in the dialog box for administrator's operation. When system occurs accident, recovery of database file can prevent loss or omission of relative information [10]. The system will show name of backup database and automatically generate the time of database backup when administrator backup data file or material. Backup of database module is good to the storage of information in the whole management system and the safety of relative data.

Service management module mainly includes service item classification, service item management, service item addition, service item review, room price addition and room booking management. Service item classification is mainly the management of product classification setting. Specific items have guest room service, catering service, entertainment and banquet service. Administrator can add, modify and delete sub-column. Jurisdiction of adding sub-column is adding subclass and choosing from the belonged large class.

4 Conclusion

Hotel management information system is related to whether the hotel can develop and the prospect of hotel industry. It is the core of long term development of hotel. Hotel industry so far devote to developing more effective management information system. It updates the management of hotel and brings rich economic benefit to hotel. With the development of network information, most industry tends to develop e-commerce. E-commerce developments of hotel industry represent the general trend. Therefore, more application of computer technology in hotel management is the develop trend of current hotel industry. Hotel computer application system is different from the computer system application of other industry. It has its own characteristic. Combining with computer technology, hotel improves the management mechanism inside hotel. It makes hotel administrators comprehensively see the business situation of hotel, meets the demand of customer's individual demand and improves service efficiency and improves economic benefit.

References

[1] Wang J 2010 Hotel management system design *Government Affairs Work* **2010**(2) 23-6

[2] Li DJ, Lu Y 2010 Research on booking function optimization of hotel portal system based on the case analysis of three kinds of hotels *Industry Economy* **2010**(12) 32-4

[3] Li XQ 2011 *Hotel information management system design based on B/S structure* Chengdu: Electronic Sci.& Tech. University **2011** 33-43

[4] Liu X 2011 *Hotel management system design and achievement based on B/S structure* Jilin: Jilin University **2011** 23-35

[5] Zhang M, Zhang GY, Ye ZL 2011 Research on the effect of online information to hotel online booking based on analysis of Xiecheng Hotel online booking data *Tourism Subject* **2011**(12) 34-5

[6] Liu T, Xie YH, Lv ZW 2012 Hotel accommodation information management system *Journal of Kiamusze University (Natural Science Edition)* **20**(4) 447-9

[7] Wang LY 2010 Realization of Excel data inquire in VB application program *Railway Computer Application* **19**(1) 50252

[8] Zhang X 2010 *Research and realization of hotel management system based on NET* Chengdu: Electronic Sci.&Tech. University **2010** 15-40

[9] He KK 2011 New progress of research on educational informationization in China *Chinese Electrochemical Education* **2011**(1) 2-3

[10] Zhang SF, Zhang J 2010 First exploration of hotel catering ordering management system *Journal of Hezhou College*

Authors

Wensu Xu, born in 1981, Hebei Province of China

Current position, grades: Lecturer
University studies: Post-Graduate degree was earned in Ocean University of China.
Scientific interest: Hotel management and tourism management

Leilei Wang, born in 1983, Shandong Province of China

Current position, grades: Lecturer
University studies: Post-Graduate degree was earned in East China Normal University.
Scientific interest: Hotel management and tourism management

Operation Research and Decision Making

Fractal model of electric conductivity of aging Al alloys

N P Pravednaya*, S F Baranova

K.D.Ushinsky South Ukrainian National Pedagogical University

Ukraine, Odessa, Staroportofrankovskaya 26, 65020

Abstract

Theoretical study of anisotropy changes of crystal structure in Al alloys as a result of senescence is performed. The appropriate model is developed that allowed to explain experimental results concerning the influence of tempering and cold rolling on the anisotropy of electrical conductivity of Al alloy D16.

Keywords: crystal anisotropy, plastic deformation, diffusion, dislocations, electrical conductivity list

1 Introduction

Aluminium (Al) alloys are the most important construction materials in the modern industry. Further development of nuclear energetics, all transport modes, space vehicles demand an application of these materials.
The typical alloying elements in these materials are copper, magnesium, manganese, silicon and zinc. There are two principal classifications of Al alloys, namely casting alloys and wrought alloys, both of which are further subdivided into the categories heat-treatable and non-heat-treatable. About 85% of aluminium is used for wrought products, for example rolled plate, foils and extrusions. Cast aluminium alloys yield cost-effective products due to the low melting point, although they generally have lower tensile strengths than wrought alloys. The most important cast Al alloy system is Al–Si, where the high levels of silicon (4.0–13%) contribute to give good casting characteristics. Al alloys are widely used in engineering structures and components where light weight or corrosion resistance is required [1, 2].

Different factors are responsible for the anisotropy of physical properties of Al alloys. Texture, or crystallographic anisotropy, is probably the most important one, but also the grain shape, precipitates and dislocation structures may influence the properties. While a single crystal is highly anisotropic, an aggregate of completely differently oriented grains might be also isotropic. This relates to the variation in strength of different orientations, which for a tensile test is given by the Schmidt factor for each grain. In textured polycrystals a large number of the grains have approximately the same orientation, causing anisotropy.

Physical characteristics of Al alloys (electric and heat conductivity, diffusion, thermal expansion) are described by second-order tensors and very structurally - sensitive [3, 4]. Studies of the anisotropy properties of Al alloys are important to understand their defect structure and their influence on service characteristics of these materials.

2 Description of the model and obtained results

As a result of plastic deformation the metals with homogeneously distributed lattice defects form a defect fractal structure [5]. It has been found that after considerable deformation the metals with cubic lattice exhibits anisotropy of physical properties that can be described by a second-order tensor.

Rectangular samples with sizes 10x100 mm were investigated. Samples were cut at various angles relative to the rolling direction (see Figure 1).

Figure 1 shows changes of electric resistivity of freshly deformed D16 alloy (as a result of quenching and rolling). One can see anisotropy of resistivity. The analogous results were observed previously in sheets of copper alloys Cu. [6].
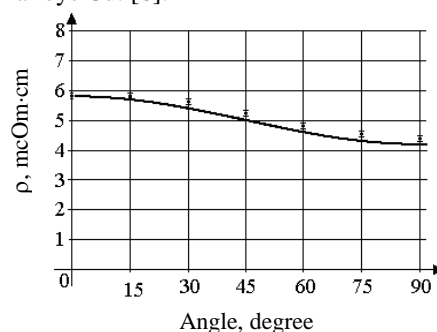

FIGURE 1 Changes of the resistivity of D16 alloys a result of quenching and rolling; the points correspond to experimental results; the solid line is obtained theoretically

In the process of significant plastic deformation (more than 50 % in thickness) as a rule, cellular dislocation structure is formed in metals and alloys. In metals with pure texture deformation cellular dislocation structure is

---
Corresponding author e-mail: 300579@mail.ru

different for different textural components. The walls of cells with large concentration of dislocations have a fractal nature with unequal fractal dimension [7 - 9].

Present study of anisotropic electric conductivity for alloy D16 is based on the accounting a fractal structure that can be depicted as in Figure 2. To compose such fractal model the details of the dislocation structure must be accounted.
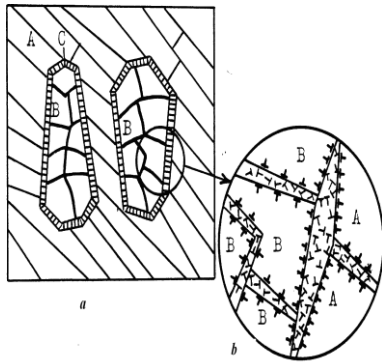

FIGURE 2
a) Fractal model of the structure of deformed sample: A, B corresponds to a highly disoriented meso- region with small intrinsic misorientation of dislocation cells, C is a boundary between strongly disoriented meso - regions (texture components) consisting of dislocations;
b) Fragment shows the accommodation between cells of meso − areas A−B (large dislocation density) and between cells A−A and B−B (low dislocation density).

It is known that aluminium has high stacking fault energy (~ 0.2 J/m^2). Therefore, a cellular dislocation structure at earlier stages of deformation is formed unlike the situation in copper, which has a lower stacking fault energy (0.04-0.05 J/m^2).

As mentioned, alloy D16 after rolling by 50% in thickness has a texture similar to the texture of the copper after rolling by 95 % in thickness. The structure of Al alloys after deformations is characterized by rectangular grains. Within the relative strain ε <1 the grain boundaries consist of a plexus of high density of dislocations. So it can be assumed that in D16 alloy the formed cellular dislocation structure similar to the structure of the deformed copper. Therefore, the model calculations of anisotropic electric conductivity for deformed alloy D16 can be based on the data of the structure of the deformed copper.

Developing the model we selected a representative volume of heterogeneous material. The inclusion matrix is a parallelepiped with edges of length l_i. The problem reduces to the calculation of the effective electric conductivity in micro-volume of homogeneous material. The details of the model are illustrated in Figure 3.

Since the orientation of sub-boundaries in general case does not coincide with the axes of the laboratory coordinate system, we introduce the crystallographic coordinate system axis Ox_1, Ox_2 and Ox_3, which are oriented along the principal axes of the electric conductivity tensor σ_{ii} (i = 1, 2, 3). Then, in accordance with the appropriate types of the dislocation structure, the axes of the electric conductivity tensor (the component

{110} <112>) and the switched (component {112} <111>) are oriented as shown in Figure 3.

Taking into account the anisotropy of the properties, the first iteration step can be written as a function of the angle φ in the plane xOy of the laboratory coordinate system:

$$\sigma(\phi)=\left[\frac{1-m}{\sigma_{22}^{M}+\left(\sigma_{11}^{M`}-\sigma_{22}^{M}\right)\cos^2\phi}+\frac{m}{\sigma_{22}^{B}+\left(\sigma_{11}^{B`}-\sigma_{22}^{B}\right)\sin^2\phi}\right]^{-1}, \quad (1)$$

where m is a volume content of inclusions (meso - region texture components {112} <111>).

$$\sigma_{11}^{M`} = \sigma_{33}^{M} + \left(\sigma_{11}^{M}-\sigma_{33}^{M}\right)\cos^2\psi_1, \quad (2)$$

$\sigma_{22}^{M`}$ is a part of the conductivity tensor matrix along $Ox_2$, which coincides with $O_y$, that is $\sigma_{22}^{M`} = \sigma_{22}^{M}$,

$$\sigma_{11}^{в`} = \sigma_{22}^{в}\sin^2\varphi_0 + \left(\sigma_{11}^{в}\cos^2\psi_2 + \sigma_{33}^{в}\sin^2\psi_2\right)\cos^2\varphi_0, \quad (3)$$

$$\sigma_{22}^{в`} = \sigma_{22}^{в}\cos^2\varphi_0 + \left(\sigma_{11}^{в}\cos^2\psi_2 + \sigma_{33}^{в}\sin^2\psi_2\right)\sin^2\varphi_0. \quad (4)$$

Then we transform the experimental values of electric conductivity in relative units as the ratio $\sigma(\varphi) = \rho_0 / \rho(\varphi)$, where ρ_0 is resistivity of the aged alloy D16 equal to 5.2 10^{-4} ohms-cm.

In the expression (1) $\sigma_{11}^{M`}$, $\sigma_{22}^{M}$, $\sigma_{22}^{в`}$ and $\sigma_{22}^{в}$ are the electric conductivity values of the matrix and inclusion, reduced to x and y axes of the laboratory system.

The calculated contribution of dislocations, in the laboratory frame, to electric conductivities relatively texture components are presented in Table 1. As we see, the values of the electric conductivity tensor expressed in the laboratory coordinate system are markedly different for the main texture components for D16 alloy. Expressions (3) - (4) show these differences.
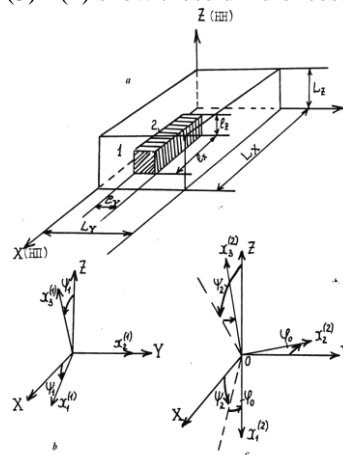

FIGURE 3 Model structure: 1 is an external matrix; 2 is an inclusion, b, c are the principal axes of the conductivity tensor matrix (texture component {110} <112>) and inclusion (texture component {112} <111>),ψ_1=30^0, ψ_2=35^0, φ_0=39^0.

TABLE 1 Dislocation contribution to electric conductance of texture components of alloy D16

Texture components	Components of the electric conductivity tensor
{110}<112>	$\sigma_{11}^{\text{м`}} = 0{,}769$   $\sigma_{22}^{\text{м`}} = 1{,}755$
{112}<111>	$\sigma_{11}^{\text{в`}} = 1{,}755$   $\sigma_{22}^{\text{в`}} = 0{,}672$

We calculate the anisotropic resistivity of D16 alloy using the data in Table 1 and experimental values of electric conductivity.

For computer simulation of the fractal model of electric conductivity of deformed alloy, we use a method of renormalization − group transformations [10]. On the first iteration step in formula (3 and 4) were substituted for the original values of the parameters of matrix and inclusion with a given volume concentration $m_1 = 0, 4$ and calculate σ_{11}^{ef} and σ_{22}^{ef} conductivity. At the second step iteration in the appropriate formula (3), (4) instead of σ_{11}^{δ} and σ_{22}^{δ} tripped conductivity coefficients σ_{11}^{ef} and σ_{22}^{δ}, calculated in the first step, and the new concentration m2, the ratio found in the interval [0, 1]:

$$m_k = m_{k-1}^2 \left(4 + 8m_{k-1} - 14m_{k-1}^2 - 40m_{k-1}^3 + 16m_{k-1}^4 + 288m_{k-1}^5 - 655m_{k-1}^6 + 672m_{k-1}^7 - 376m_{k-1}^8 + 112m_{k-1}^9 - 14m_{k-1}^{10} \right) \quad , (5)$$

The function (5) is shown in Figure 4.

As a result equation (1) gives a curve of the ellipse form relatively the principal axes. Thus, we obtained that the electric conductivity of deformed Al alloy D16 with a cubic lattices described by a second-order tensor.

FIGURE 4 The function $m_k = f\left(m_{k-1}\right)$ determined at the interval [0, 1]. Solid line reflects the change of dislocation concentration and their inclusion in the calculation σ_{11}^{ef} and σ_{22}^{ef}.

3 Conclusions

1) The anisotropy of electric conductivity was measured for D16 Al alloy with cube lattice after tempering and cold rolling.

2) It was established that the electric conductivity of studied alloy in the direction of rolling is 30 % less than in transversal direction of sheet. The anisotropy of electric conductivity can be described by the second-order tensor, which is consistent with the fractal model and experimental results.

3) Using a developed fractal model of the structure of deformed metallic materials it was determined the dislocation role in formation of the main components of texture that explain the nature of the general anisotropy of electric conductivity of alloy D 16.

Acknowledgments

References

[1] King F 1987 *Aluminum and its Alloys*. *Ellis Harwood Series in Metals Materials* Chichester England: Ellis Harwood
[2] Sanders R E 2001 *The Journal of the Minerals* **53**(2) 21–5
[3] Williams J C, Starke E A Jr 2003 *Acta Materialia* **51** 5775–99
[4] Degarmo E P, Black J T, Kosher R A 2003 *Materials and Processes in Manufacturing*: Wiley
[5] Hähner P, Bay K, Zaiser M 1999, *Acta Materialia* **47** 2463–76
[6] Barlat F, Liu J 1998 *Mater. Sci. Eng*. A **257** 47–61

[7] Randle V, Engler O 2000 *Introduction to Texture Analysis* NY: Gordon and Breach Science Publishers
[8] Charles A Harper1993 *Electronic Materials and Processes - Handbook, Third Edition*: McGraw-Hill HANDBOOKS
[9] Voss R F, R B Laibowitz, and Allessandrini E I 1982 *Phys Rev Lett* **49** 1441–5
[10] Newman M E J, Watts D J 1999 *Phys Lett* A **263** 341-6

Authors

Nadejda Pravednaya

Current position: Assistant Professor, PhD, Department of Physical and Mathematical Modelling, South-Ukrainian National Pedagogical University
Scientific interest: Real structure, mechanical properties of metals and alloys
Publications: 30
Experience: X-ray structural analysis of metallic materials, modelling of defect structure and phase transformations in metals and alloys

Svetlana Baranova

Current position: M Sci student, Department of Physical and Mathematical Modelling, South-Ukrainian National Pedagogical University
Scientific interest: Mechanical properties of metals and alloys
Publications: 2
Experience: X-ray structural analysis of metallic materials.

The finite time thermodynamics analysis and the energy-saving optimization of the coil organic heat transfer material heater

Wei Li Gu[1, 2]*, Han Qing Wang[3], Guang Xiao Kou[3], Qiao Ying Cao[2]

[1]*School of Energy and Power Engineering, Central South University, Changsha 410000, China*

[2]*Department of Urban Construction, University of South China, Hengyang 421001, China*

[3]*Hunan University of Technology, Zhuzhou 412000, China*

Abstract

According to the problems as low efficiency, aging of organic heat transfer material and frequent accidents in the operation of the coil organic heat transfer material heater, with the finite time thermodynamics, this paper studies the actual processes including combustion process, flow and heat transfer process, considers the irreversible factors as combustion condition, the fuel characteristics, flow resistance and heat transfer temperature difference, derives the calculation formula of energy loss for three processes and to determine the minimum energy loss. Based on this, this paper proposes energy-saving optimization strategy and gives examples, the results accord with the analysis. In design and operation, the outlet flue gas temperature of furnace, flue gas temperature and flow velocity of hot oil are important.

Keywords: the coil organic heat transfer material heater, finite time thermodynamics, energy loss, energy-saving optimization

1 Introduction

Nowadays, the coil organic heat transfer material heaters are widely used in industrial processes in which indirect heating under high temperature is needed. Compared with traditional vapour heating, the organic heat transfer material heating has many advantages such as high temperature under low pressure, without heat loss of condensation, high efficiency of heating system, but the most important is the energy-saving effect by 35~45% theoretically. However, in the operation of the coil organic heat transfer material heater, there are still some problems as low efficiency, aging of organic heat transfer material, frequent accidents, and so on. For the heater, there exist three processes, which occur at the same time. These processes are the combustion process of fuel, the flow process of oil and the heat transfer process between the flue gas and oil. These processes are irreversible with the irreversible factors such as the fuel, combustion condition, flow resistance and heat transfer temperature difference. The more irreversible of the process, the great the entropy generation, and the great the energy loss, this means the waste of energy. Thus, in this paper, the energy loss in the coil organic heat transfer material heater is analysed, the measures to reduce the energy loss are studied, and furthermore the energy-saving optimization strategy is obtained.

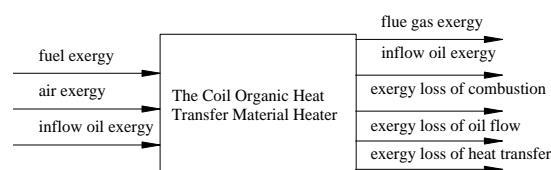In the operation of the coil organic heat transfer material heater, the energy flow is shown in Figure 1.

FIGURE 1 The energy flow in the coil organic heat transfer material heater

2 The finite time thermodynamics analysis of combustion process

The combustion process of fuel is an intensive chemical reaction, and the reaction heat released depends on the types of fuel and chemical reaction conditions directly. According to the thermodynamics principles, the combustion is irreversible. For this process, the entropy generation indicates the irreversible degree of the process, and the energy loss indicates the energy loss in the process. The more irreversible the process, the greater the entropy generated, and the greater the energy losses.

Combustion process occurs in the furnace, fuel and air entering the furnace, mixing and burning, and producing flue gas. The furnace is a control volume system, the media in and out the system are fuel, air and flue gas. However, as the energy flow of the furnace, besides the energy of fuel, air and flue gas, there exists energy loss.

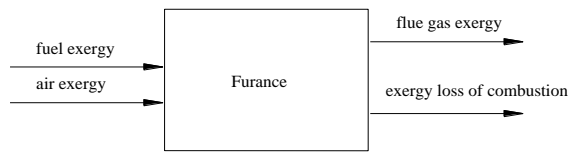
Corresponding author e-mail: guweili@126.com

FIGURE 2 Energy flow of adiabatic combustion in furnace

Assume the combustion process is adiabatic, the energy flow of the process is shown in Figure 2. The energy entering the system are the fuel energy and the air energy, the energy leaving the system is the flue gas energy, besides, there exists energy loss caused by the irreversibility of combustion and the energy balance equation of combustion is as following [1].

$$E_{x,f} + E_{x,a} = E_{x,g} + E_{x,l} . \tag{1}$$

From Equation (1), the energy loss can be calculated as following:

$$E_{x,l} = E_{x,f} + E_{x,a} - E_{x,g} . \tag{2}$$

The energy loss can be obtained by determining the other terms in the energy balance equation.

2.1 FUEL ENERGY

The fuel energy is the energy entering the system with fuel, which consists of chemical energy $\left(E_{x,f,ch}\right)$ and physical energy $\left(E_{x,f,ph}\right)$. Because $E_{x,f,ch} \gg E_{x,f,ph}$, the $E_{x,f} \approx E_{x,f,ch}$. For chemical energy, the standard state is 1 atm, 25°C. For gas fuel, the chemical energy can be obtained with theory formula, and for solid and liquid fuel, the experienced formulas are used to obtain the chemical energy because of their complex component.

2.1.1 Gas fuel energy

The standard chemical energy of gas fuel can be obtained with following Equation (2), (3):

$$E_{x,f}^0 = -\Delta G_n^0 + \sum n_j E_{x,j}^0 - \sum n_{O_2} E_{x,O_2}^0 = \\ -(\Delta H_f^0 - T_0 \Delta S_f^0) + \sum n_j E_{x,j}^0 - \sum n_{O_2} E_{x,O_2}^0 . \tag{3}$$

When the environmental state is different with standard state, the temperature must be corrected.

2.1.2 Liquid fuel energy and solid fuel energy

The standard chemical energy of solid fuel and liquid fuel can be obtained with following Equation:

$$E_{x,f}^0 = 34218.87[C] + 21.97[N] + 116702.76[H] + 18260.357[S] - 13278.593[O] + 24114.107[F] + 11759.425[Cl] + \\ 5038.791[Br] + 2897.42[I] - (298.15)(0.71768)m_{a,h} + 0.6276[O] + 32792.8[C] + 141791.11[H] + 16019.49[S] - \tag{4} \\ 17723.842[O] + 18607.798[F] + 7493.126[Cl] + 2853.363[Br] + 1401.054[I].$$

2.2 FLUE GAS ENERGY

The flue gas energy consists of chemical energy and physical energy. The chemical energy can be omitted because of the fixed components of flue gas in engineering applications. And, for flue gas, the physical energy is dominant. The standard chemical energy of solid fuel and liquid fuel can be obtained with following equations:

$$E_{x,ph}^g = \sum x_i [(H_g - H_g^0) - T_0(S_g - S_g^0)] , \tag{5}$$

$$E_{x,ph}^g = \int \sum x_i c_{pi} dT - T_0 \int \sum x_i c_{pi} \frac{dT}{T} . \tag{6}$$

2.3 ENERGY LOSS IN COMBUSTION

From Equation (2), the energy loss can be obtained with fuel energy, air energy and flue gas energy.

2.4 THE ENERGY LOSS OF TYPICAL FUELS

2.4.1 Basic assumptions

For simplification, consider following assumptions: the quantities of solid fuel, liquid fuel and gas fuel are 1 kg, 1 kg, and 1 m³; the environmental state is standard state, so the air energy is zero; the combustion is adiabatic, the temperature is theoretical combustion absolute temperature; the combustion is completed combustion without excess air coefficient; the function of specific heat capacity is [2]:

$$c_{pm} = a_0 + a_1 \times 10^{-3} T + a_2 \times 10^{-6} T^2 . \tag{7}$$

2.4.2 Typical fuels

The solid fuel is bituminous coal (AII), theoretical combustion absolute temperature is 1767 K/1494℃, the flue gas volume is 5.43 m³ with 1 kg fuel, $Q_{net,v,ar}$ is 18726.4 kJ/kg.

The characteristics of bituminous coal (AII) are shown in Table 1.

TABLE 1 Matter fraction of solid fuel (%)

C_{ar}	H_{ar}	O_{ar}	N_{ar}	S_{ar}	W_{ar}	A_{ar}	V_{daf}
47.53	3.56	6.95	0.9	0.33	10.33	30.4	28.8

The liquid fuel is light diesel oil (0#), theoretical combustion absolute temperature 2128 K/1855°C the flue gas volume is 11.587 m³ with 1kg fuel, $Q_{net,v,ar}$ is 42914.7 kJ/kg.

The characteristics of light diesel oil (0#), are shown in Table 2.

TABLE 2 Matter fraction of liquid fuel (%)

C_{ar}	H_{ar}	O_{ar}	N_{ar}	S_{ar}	W_{ar}	A_{ar}
85.55	13.49	0.66	0.04	0.25	0	0.01

The gas fuel is pyrogenous coal gas, theoretical combustion absolute temperature 1959 K/1868°C, the flue gas volume is 4.28 m³ with 1 m³ fuel. The function of specific heat capacity is:

$$c_{pm} = a_0 + a_1 \times 10^{-2} T. \tag{8}$$

The characteristics of pyrogenous coal gas are shown in Table 3.

TABLE 3 Volume fraction of gas fuel (%)

	H_2	N_2	CH_4	NH_3
x_i	66.0	21.0	9.0	4.0
a_0	29.08	28.882	19.874	27.55
a_1	-0.1918	-0.1570	5.0242	2.5644

2.4.3 The energy loss of typical fuels

With the data and formula above-mentioned, the energy loss of combustion with typical fuels is shown in Table 4.

TABLE 4 Energy loss of combustion with different fuels

	Solid Fuel	Liquid Fuel	Gas fuel
Fuel energy	20339.65 kJ/kg	45427.1 kJ/kg	10892.9 kJ/m³
Flue gas energy	9130.68 kJ/kg	27628.87 kJ/kg	9103.37 kJ/m³
Energy loss	11208.97 kJ/kg	17798.23 kJ/kg	1789.53 kJ/m³
Energy loss ratio	55.1%	39.2%	16.4%

2.5 NOTATION

The following terms are used in equations above-mentioned:

$E_{x,f}$ – fuel energy;

$E_{x,a}$ – air energy;

$E_{x,g}$ – fuel gas energy;

$E_{x,l}$ – energy loss;

$E_{x,f,ch}$ – chemical energy;

$E_{x,f,ph}$ – physical energy;

$E_{x,f}^0$ – standard chemical energy of gas fuel;

ΔG_n^0 – Gibbs function;

$n_j, E_{x,j}^0$ – the mole number and chemical energy of resultant according to 1 kmol fuel;

n_{O_2}, E_{x,O_2}^0 – the mole number and chemical energy of O_2 needed for 1 kmol fuel;

ΔH_f^0 – standard reaction enthalpy;

x – the matter fraction of elements;

$m_{a,h}$ – the matter fraction of ash;

T – combustion temperature;

T_0 – environment temperature;

c_{pm} – specific heat capacity;

a_0 and a_1 – constants.

2.6 ANALYSIS OF ENERGY LOSS WITH THREE TYPICAL FUELS

For gas fuel, liquid fuel and solid fuel, the energy loss ratio is 16.4%, 39.2% and 55.1%. The energy loss is relatively great in combustion, and the energy utilization ratio is relatively low. The effect factors include the combustion condition and the fuel characteristics.

2.6.1 The effect of combustion condition-temperature

As shown in Equation (2), under certain environmental state, the fuel energy is constant, and the energy depends on the flue gas energy directly. As flue gas energy increases, the energy loss reduces. From Equations (5) and (6), flue gas energy depends on the combustion temperature. The higher the temperature, the great the flue gas energy. The combustion temperature of solid fuel is lower than that of the liquid fuel and gas fuel in the example, and the energy loss ratio is greatest.

2.6.2 The effect of fuel characteristics-component

The flue gas energy also depends on the quantity of flue gas produced by combustion, and the quantity of flue gas depends on the component of fuel directly. The solid fuel and liquid fuel has the same elements, but different matter fraction, especially on the C and H, which results in the different quantity of flue gas. In this example, the matter fraction of C is 85.55% and 47.53%, in liquid fuel and solid fuel, the matter fraction of H is 13.49% and 3.56%, thus the energy loss ratio of solid fuel is higher than that of liquid fuel.

2.6.3 The effect of fuel characteristics-existing form of flammable substances in fuel

In the combustion, the flammable substances in solid fuel and liquid fuel must separates out firstly, which causes excess energy consumption, thus the energy loss ratio of solid fuel and liquid fuel is higher than that of gas fuel.

3 The finite time thermodynamics analysis of flow and heat transfer process

In the operation of the coil, organic heat transfer material heater, the hot oil flows in tubes and absorbs heat. When

the temperature of hot oil increases to specified temperature, the hot oil leaves boiler and enters the heat consumers. That means the flow process and heat transfer process occurs at the same time. The model of the flow and heat transfer process of hot oil in tubes is shown in Figure 3.
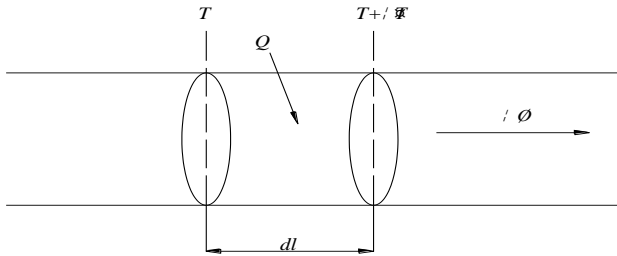

FIGURE 3 The flow and heat transfer process of hot oil in tube

Figure 3 shows the flow and heat transfer process of m kg hot oil in tube. Set Infinitesimal length be dl, set the heat absorbed per unit tube length be Q, the temperature of hot oil increases from T to $T+\Delta T$, the velocity of hot oil in the tube be ω. Because of the flow resistance and the temperature difference in heat transfer, the flow and heat transfer process is irreversible, and the entropy generation can be obtained with follows:

$$dS_g = mds - \frac{Qdl}{T+\Delta T}.$$ (9)

According to the thermodynamic relations, considers the relationship between enthalpy change and heat in isobaric process, obtains the entropy change in the process shown in Equation (10).

$$ds = \left(\frac{Q}{m} - \frac{1}{\rho} \frac{dp}{dl} \right) \frac{dl}{T}.$$ (10)

Combined Equations (9) and (10), obtains the entropy generation per unit length shown in Equation (11).

$$S_g = \frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right) + \frac{Q\Delta T}{T^2 \left(1+\frac{\Delta T}{T} \right)}.$$ (11)

Considers time factor τ, obtains the entropy generation ratio per unit length shown in Equation (12).

$$S_g^{\tau} = \frac{\frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right) + \frac{Q\Delta T}{T^2 \left(1+\frac{\Delta T}{T} \right)}}{\tau}.$$ (12)

When $\tau \to 0$, the entropy generation ratio per unit length is shown in Equation (13).

$$S_g^{\tau} = \frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right) + \frac{Q\Delta T}{T^2}.$$ (13)

The energy loss ratio per unit length is shown in Equation (14).

$$L = S_g^{\tau} T_0 = \left[\frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right) + \frac{Q\Delta T}{T^2} \right] T_0.$$ (14)

From Equation (13), the entropy generation ratio consists of two parts, the first part is entropy generation ratio of dissipation effect which caused by the flow resistance, the second part is entropy generation ratio of potential difference which caused by the temperature difference in heat transfer. The flow resistance and temperature difference are the irreversible factors of the flow and heat transfer process. The entropy generation ratio of dissipation effect depends on the type of hot oil, the flow pattern, physical parameters, structure and operation of the coil organic heat transfer material heater. The entropy generation ratio of heat transfer with temperature difference depends on heat transfer coefficient, heat transfer area, mean exothermic temperature and mean endotherm temperature.

3.1 THE ENERGY LOSS RATE OF DISSIPATION EFFECT

From Equation (14), the energy loss ratio can be obtained based on the entropy generation ratio and the environment temperature. From Equation (13), obtains the entropy generation ratio of dissipation effect shown in Equation (15).

$$S_g^{\tau} = \frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right).$$ (15)

From fluid mechanics, the entropy generation ratio and energy loss ratio of dissipation effect are related to the on-way resistance λ, Reynold number Re, viscosity v, velocity ω and pipe diameter d, as shown in Equations (16) and (17).

$$S_g^{\tau} = \frac{m}{\rho T} \left(-\frac{\Delta p}{l} \right),$$ (16)

$$L = S_{g,f}^{\tau} T_0 = \frac{\lambda}{d} \frac{\varpi^2}{2T} T_0 = \frac{\lambda v^2 \operatorname{Re}^2}{2Td^3} T_0.$$ (17)

From Equation (17), energy loss ratio L is the function of λ, ω, d and T. Among these parameters, λ depends on the Re, and the Re is related to the v, ω and d. In engineering, when a heater puts into practice, the pipe diameter is determined, that means the energy loss ratio of dissipation effect mainly depends on the viscosity and velocity. Viscosity v depends on the type of hot oil. The influence of temperature is on the viscosity. So the comprehensive influence of viscosity and velocity can be considered with constant pipe diameter.

The following example is a comparison among three types of hot oil (YD-300, XD-300, L-Q300), calculation conditions: L=1 m, d=45 mm, k=0.046 mm. The example is about the influence of viscosity and velocity on the energy loss ratio. The results are shown in Figures 4-6.
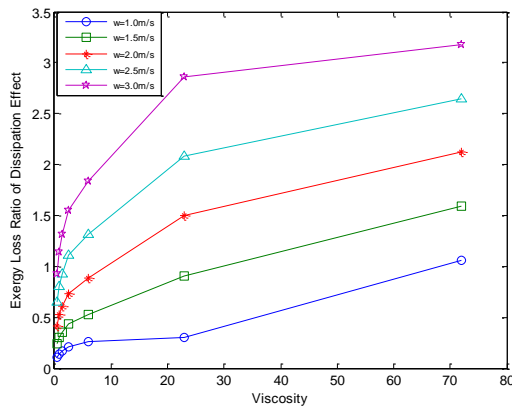
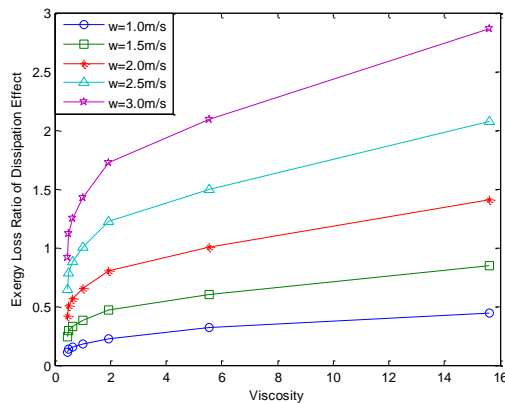FIGURE 4 The influence of viscosity and velocity for XD-300

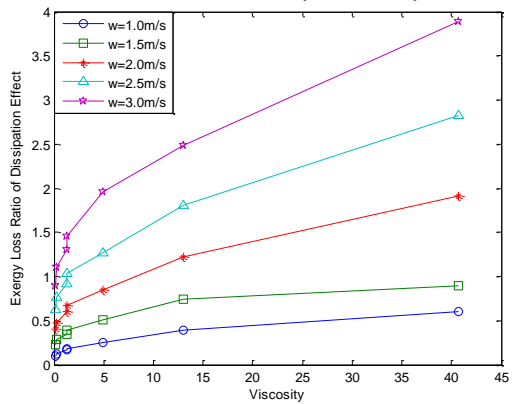FIGURE 5 The influence of viscosity and velocity for YD-300

FIGURE 6 The influence of viscosity and velocity for L-Q300

Figures 4-6 show the influence of viscosity and velocity on energy loss ratio of dissipation effect. The influence is similar for the three types of hot oil. With the increase of temperature, viscosity decreases, and the energy loss ratio of dissipation effect decreases, on the other hand, with the increase of velocity, the energy loss ratio of dissipation effect increases. In operation, the organic hear transfer material experiences temperature rising stage and normal operation stage. In the temperature rising stage, because of the great variation of hot oil viscosity, the influence of viscosity on energy loss ratio of dissipation effect is larger than that of the velocity. However, in the normal operation stage, the viscosity is basically steady, and the influence of velocity is dominant. For different type of hot oil, the variation range is different,

so the temperature range in which the viscosity and velocity is dominant in part is different. Thus, in operation, the type of hot oil must be considered to decrease the energy loss ratio of dissipation effect and the velocity must be controlled to avoid overheat.

3.2 THE ENERGY LOSS RATE OF HEAT TRANSFER WITH TEMPERATURE DIFFERENCE

The hot oil absorbs heat from flue gas in the coil organic heat transfer material heater. To simplify the analysis, mean temperature is adopted, high temperature flue gas releases heat at mean exothermic temperature, and hot oil absorbs heat at mean endotherm temperature. The heat released is equal to that absorbed.

According to the thermodynamic theory, the energy loss ratio of heat transfer with temperature difference in isolated system is as follows [5, 6]:

$$L^\tau = T_0 \Delta S_{iso}^\tau = \frac{T_0 \Delta S_{iso}}{\tau} = \frac{T_0 Q \left(\dfrac{1}{T_{Lm}} - \dfrac{1}{T_{Hm}} \right)}{\tau} = \tag{18}$$
$$K T_0 \left(T_{Hm} - T_{Lm} \right) \left(\frac{T_{Hm} - T_{Lm}}{T_{Hm} T_{Lm}} \right).$$

From Equation (18), as the environment temperature is determined, the energy loss ratio mainly depends on heat transfer coefficient, heat transfer area, mean exothermic temperature and mean endotherm temperature. When the structure and medium of the organic heat transfer material heater are determined, the energy loss ratio only depends on mean exothermic temperature and mean endotherm temperature. To obtain smallest energy loss, the mean endotherm temperature must be near to the mean exothermic temperature, thus the irreversible degree of heat transfer with temperature difference is small, and so does the energy loss. The following is an example of energy loss ratio calculation, the type of the organic heat transfer material heater is QXS0.93-280/260-Y. The fuel is 0# light diesel oil, and different mean exothermic temperature and mean endotherm temperature are adopted. The results are shown in Table 5.

TABLE 5 Energy loss of heat transfer with temperature difference

Name	Symbol	Unit	Value 1	Value 2
Flue gas temperature	Θ_{py}	°C	380	300
The outlet flue gas temperature of furnace	Θ_{ll}	°C	938	800
The inlet temperature of heat transfer oil	T_{l1}	°C	260	260
The outlet temperature of heat transfer oil	T_{l2}	°C	280	280
Low thermal value	Q_{dw}^y	LJ/kg	42914.7	42914.7
The mean endotherm temperature	T_{lm}	°C	270	270
The mean exothermic temperature	T_{hm}	°C	618	486
Energy efficiency	$\eta_{ex.B}$	/	32.9	34.9
Energy loss ratio	L		67.1	65.1

3.3 NOTATION

The following terms are used in equations above-mentioned:

dl – Infinitesimal length
Q – the heat absorbed per unit tube length
T – temperature of hot oil
ΔT – temperature increase
ω – velocity of hot oil in the tube
s – entropy generation
L – energy loss ratio
λ – on-way resistance
Re – Renold number
$\boldsymbol{v}$ – viscosity
d – pipe diameter
k – heat transfer coefficient
F – heat transfer area
T_{Hm} – mean exothermic temperature
T_{Lm} – mean endotherm temperature

4 The energy-saving optimization strategy of the organic heat transfer material heater based on finite time thermodynamic analysis

Based on the finite time thermodynamic analysis on the irreversible processes, the main influencing factors of combustion process, flow process and heat transfer process are obtained. By analysing these factors, the energy-saving optimization strategy can be achieved in design and operation management.

4.1 DESIGN

4.1.1 Determination of Design Parameters

In the design of the coil organic heat transfer material heater, the design parameters related to combustion and heat transfer with temperature difference are the outlet flue gas temperature of furnace and flue gas temperature. The design parameters related to flow process is velocity of hot oil.

1) Determination of the outlet flue gas temperature of furnace. The outlet flue gas temperature of furnace is a parameter reflecting the heat operation capacity in furnace. Too lower the outlet flue gas temperature of furnace can cause the lower reaction temperature, which causes less flue gas energy, increases the energy loss and energy loss ratio of combustion, and reduces the energy utilization efficient. However, the reduction of the outlet flue gas temperature of furnace can reduce the mean exothermic temperature, decrease the energy loss of heat transfer process with large temperature difference, and improve the transfer efficiency of energy. So in order to determine an appropriate outlet flue gas temperature of furnace, it is necessary to consider the two aspects in design. However, in actual design process, the determination of the outlet flue gas temperature of furnace is experiential, and lacks theoretical basis. Thus sets energy loss as evaluation index, obtains function relationship between the total energy loss of irreversible processes and the outlet flue gas temperature of furnace. Setting this functional relationship as objective function, the outlet flue gas temperature of furnace can be optimized, and provides theoretical basis for the determination of the outlet flue gas temperature of furnace.

2) Determination of flue gas temperature. The flue gas temperature is the temperature of flue gas emitted into the environment. The flue gas temperature affects the energy loss of heat transfer with temperature difference. The higher flue gas temperature means the higher mean exothermic temperature, and the more the energy loss; the lower flue gas temperature means the lower mean exothermic temperature, and the less the energy loss. However the flue gas temperature must be higher than the outlet temperature of heat transfer oil. The temperature difference determines the area of heating surface and affects the cost of equipment directly. However, in actual design process, the determination of the flue gas temperature is experiential, and lacks theoretical basis. Thus sets energy loss as evaluation index, establishes function relationship between the total energy loss of irreversible processes and the flue gas temperature. Setting this functional relationship as objective function, the flue gas temperature can be optimized, and provides theoretical basis for the determination of the flue gas temperature.

3) Determination of flow velocity of hot oil. In determining the flow velocity of hot oil, two aspects must be considered. One is the energy loss caused by the higher velocity, the other is the overheat of hot oil caused by the lower velocity. Besides, the higher velocity means large circulation pump, which causes further waste of energy. Thus, the determination of flow velocity must base on the flow pattern of the hot oil in tubes. The appropriate velocity means that the flow pattern of hot oil in tubes is turbulence without too large Re.

4.1.2 Design of structure

In design of structure, the related parts include the determination of the ratio between radiant heating surface and convective heating surface, and arrangement of assistant heating surface. Based on the analysis above-mentioned, the ratio between radiant heating surface and convective heating surface can be described with the outlet flue gas temperature of furnace. Besides, the component of fuel affects the capacity of flue gas energy, and the reaction temperature in furnace affects the production of flue gas. Thus, the optimization on the ratio between radiant heating surface and convective heating surface can be achieved at the basic of the optimization on the outlet flue gas temperature of furnace with the characteristics of fuels. Because higher flue gas temperature can cause larger energy loss, the waste heat recovery devices should be

adopted for flue gas. The increased cost of devices can be evaluated with economic analysis.

4.2 OPERATION MANAGEMENT

Based on the analysis above-mentioned, the optimization on the operation management can be carried out from following aspects:

1) Implementing the real-time monitoring on the outlet flue gas temperature of furnace and the flue gas temperature in order to avoid the increase of energy loss caused by large temperature difference in heat transfer.
2) Prohibiting the change of design fuel optionally in order to avoid the increase of energy loss of combustion caused by the change of flue gas energy with different fuels.
3) Implementing the real-time monitoring on the combustion in order to avoid the increase of energy loss of combustion caused by the combustion instability under lower reaction temperature.
4) Implementing the real-time monitoring on flow pattern to avoid the increase of energy loss of flow process caused by the higher velocity of hot oil, besides avoid the overheat of hot oil at too lower velocity.

5 Conclusions

1) Studies the combustion process based on the theory of finite time thermodynamics, compares the energy loss of different fuels, analyses the influencing factors on energy loss of combustion.
2) Studies the flow process and heat transfer process based on the theory of finite time thermodynamics, compares energy loss at different mean exothermic temperature and mean endotherm temperature, compares the energy loss of different types of hot oil, analyses influencing factors of the flow process and heat transfer process.
3) On the basis of this analysis, proposes the energy-saving optimization measures on design and operation management of the coil organic heat transfer material heater, and specially points out that in the design process, objective function can be constructed with the energy loss as evaluation index to determine the outlet flue gas temperature of furnace and the flue gas temperature, and provides theoretical basis for the determination of design parameters.

Acknowledgments

References

[1] Lian L M 2007 Engineering Thermodynamics fifth edition *Beijing: China Building Industry Press (in Chinese)*
[2] Zhu M S 1988 Energy Analysis on Energy System *Beijing: Tsinghua University Press (in Chinese)*
[3] Xiang X Y 1988 Engineering Energy Analysis *Beijing: Petroleum Industry Press (in Chinese)*
[4] Wang Y 2008 Research of Irreversible Loss in Flow and Mixed Process *Master Thesis* Harbin Institute of Technology Harbin *(in Chinese)*
[5] Fu Q S and Feng X 2001 Inevitable energy loss in heat engine cycle, in Magnetism *Journal of Xi'an Jiaotong University* 35 1105-8 *(in Chinese)*
[6] Yang S M 2006 Heat Transfer fourth edition *Beijing: Higher Education Press (in Chinese)*

Authors

Weili Gu, born in October, 1974, Xinjiang Province, P.R. China

Current position, grades: University of South China, associate professor.
University studies: Central South University.
Scientific interest: industrial engineering, device simulation.
Publications: 20.

Hanqing Wang, born on October 26, 1963, Hunan Province, P.R. China

Current position, grades: Hunan University of Technology, professor.
University studies: PhD degree in Heat, Gas Supply, Ventilation and Air Conditioning Engineering, Hunan University, 2003.
Scientific interests: CFD-simulation on IAQ, building environment and facilities energy saving.
Publications: 200.

Guangxiao Kou, born on October26, 1963, Shanxi Province, P.R. China

Current position, grades: professor.
University studies: Central South University.
Scientific interests: building-saving technology; application of thermal fluid heater energy saving.
Publications: 56.

Qiaoying Cao, born in February 1989, Hunan Province, P.R. China

Current position, grades: Postgraduate in Heat, Gas Supply, Ventilation and Air Conditioning Engineering, University of South China.
University studies: University of South China (school of urban construction) 2009.
Scientific interest: industrial engineering, device simulation.

Plastic coefficient on-line calculation method for hot rolling

Chunyu He*, Zhijie Jiao, Di Wu

[1]*The State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110004, China*

Abstract

Aiming at the requirement of the high precision rolled piece plastic coefficient in the hot rolling process, this paper puts forward an on-line calculation method for the plastic coefficient. Based on the rolling mechanism model and fitting the plastic curve by the quadratic curve, the paper is to solve the tangency at the rolling pressure point, which is the plastic coefficient. The plastic coefficient calculation method exploited in the paper could be embedded in the basic automation directly before the calculation of the automatic gauge control (AGC) regulation quantity. According to the changes of the rolling force and the roll gap, the method can make the correction computation on the plastic coefficient, so as to improve the thickness compensation accuracy of the AGC system.

Keywords: hot rolling, plastic coefficient, quadratic curve, AGC

1 Introduction

At present, the hot rolling mills generally adopt the automatic AGC. The common AGC control models are the gauge meter AGC control model and the dynamic setting AGC model [1-4], the forms of these models are shown as Equations (1) and (2):

$$\Delta S_k = -\frac{M+Q}{M} \cdot \Delta h_k, \tag{1}$$

$$\Delta S_k = -\left(\frac{Q}{M} \cdot \Delta S_{k-1} + \frac{M+Q}{M^2} \cdot \Delta P_k\right), \tag{2}$$

where, M is the mill stiffness, Q is the rolled piece plastic coefficient, ΔS_{k-1} is the roll gap control quantity at the last time.

In the AGC control model, the mill stiffness M and the rolled piece plastic coefficient Q are the key parameters and they directly affect the rolled piece thickness compensation accuracy. Through the spring curve obtained during the stiffness test, the mill stiffness based on strip width compensation gets the accurate calculation. The plastic coefficient Q is defined as the needed rolling force when the rolled pieces come up with the unit transformation [5-7], that is:

$$Q = -\frac{\partial P}{\partial h}, \tag{3}$$

where, Q is the parameter to reflect the rolled pieces' mechanical properties in rolling process. In the present AGC system, the plastic coefficient calculation is mainly accomplished by the process computers. According to the rolling schedule, it needs to calculate out an approximate plastic coefficient by the equation $P/(H-h)$ as the each pass,

and send to basic automation as AGC control parameter before biting rolled piece [8-11].

In mill spring and plastic curve diagram, the average plastic coefficient is compared with actual plastic coefficient, as shown in Figure 1. The intersection point of the plastic curve and the coordinate axis h corresponds to the entry thickness of rolled pieces. The intersection point of plastic curve and elastic curve corresponds to the exit thickness of rolled pieces. The tangent slope of the point on the plastic curve is actual plastic coefficient. During the rolling process, the shape of plastic curve will change because of the plate width fluctuation, the uneven temperature of longitudinal direction, and other conditions. So the exit thickness and plastic coefficient is constantly changing. This is the reason for on-line calculation of plastic coefficient. Average plastic coefficient is directly calculated using the average entry thickness, exit thickness and average rolling force. There are following drawbacks in adopting the method of average calculation.
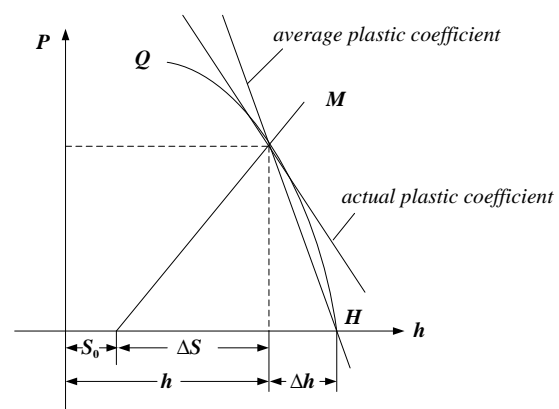
H – entry thickness, *h* – exit thickness, *S₀* – roll gap, *ΔS* – mill spring

FIGURE 1 The comparison of average plastic coefficient and real value

---

Corresponding author e-mail: hecy@ral.neu.edu.cn

The error of the approximate calculation is large, namely the difference between the tangency of actual rolling pressure point and the average slope is large.

During the rolling process, because the rolled piece is affected by the temperature difference between the head and the tail, the watermark and other factors, the plate thickness fluctuation directly affects the actual value of the plastic coefficient. So it will cause the system thickness control to be inaccurate in using a fixed average plastic coefficient to conduct AGC model calculation.

The calculation of average plastic coefficient is often done by process computer, so the control functions of automation system depend on the process computer. When the process computer is down, the accuracy of thickness control will be lower.

2 Model building

According to the calculation and application of plasticity coefficient, this article proposes a calculation method using a quadratic curve fitting plastic curve, so that the tangent slope at the pressure position can be easily solved. The method for solving the plastic coefficient does not depend on the process computer, it can be directly embedded in the basic automation and calculated in each PLC cycle time. With the rolling processing, according to the change of the rolling force and roll gap, the plasticity coefficient continues to be modified. Thereby, the compensation precision of thickness control system is improved.

The force calculation equation in hot rolling process adopts the SIMS equation [12, 13]:

$$P = 1.15 \cdot \sigma \cdot Q_P \cdot l_C \cdot B \,, \tag{4}$$

where, σ is the resistance to deformation, Q_P is the shape influence function in the deformation zone, l_C is the contacted arc length considering of the elastic flattening, B is the rolled piece's width.

The resistance to deformation model is defined in the following equation:

$$\sigma = a_0 \cdot \exp(a_1 + a_2 T) \cdot \varepsilon^{a_3} \cdot \dot{\varepsilon}^{a_4} \,, \tag{5}$$

where, $T = (t^\circ + 273)/1000$, ε is the strain, $\dot{\varepsilon}$ is the strain rate, and $a_0 \sim a_4$ are the regression coefficients corresponding to different steel grade.

The hot rolling process is generally multi-passes rolling. For the first pass, the entry thickness is the slab thickness and for other passes, it adopts the Equation (6) to calculate the entry thickness:

$$H = gap_{avg} + (P_{avg} - P_{zero})/M + W \,, \tag{6}$$

where, gap_{avg} is the average gap of the last pass, P_{avg} is the average rolling force of the last pass, P_{zero} is the zero point of rolling force, M is the mill stiffness considering the compensation of the roll size and the rolled pieces' width, W is other thickness influence items.

The actual exit thickness of rolled pieces can be calculated according to date from the production site, as shown in Equation (7).

$$h_{act} = gap_{act} + (P_{act} - P_{zero})/M + W \,, \tag{7}$$

where, h_{act} is the rolled pieces' actual thickness in the rolling process, gap_{act} is the actual gap; P_{act} is the actual rolling force.

The plastic curve reflects the relationships between rolled pieces' thickness and rolling force. In order to use conic to fit plastic curve, we need to solve out the key points in the plastic curve. There are two known key points: $(H, 0)$, and (h_{act}, P_{act}).

Suppose that in the rolling process, the strain is ε, according to Equation (4), we can get the rolling force $P(\alpha\varepsilon)$ when strain is $\alpha\varepsilon$ ($0 < \alpha < 1$). The ratio of $P(\varepsilon)$ to $P(\alpha\varepsilon)$ is shown as following formula:

$$\frac{P(\varepsilon)}{P(\alpha\varepsilon)} = \frac{\sigma(\varepsilon)Q_P(\varepsilon)l_c(\varepsilon)B(\varepsilon)}{\sigma(\alpha\varepsilon)Q_P(\alpha\varepsilon)l_c(\alpha\varepsilon)B(\alpha\varepsilon)} \,. \tag{8}$$

In the rolling process, the changes of deformation zone shape influence function Q_p and rolled piece width B is small, they can be ignored. The Equation (8) can be simplified as followed:

$$\frac{P(\varepsilon)}{P(\alpha\varepsilon)} = \frac{\sigma(\varepsilon)l_c(\varepsilon)}{\sigma(\alpha\varepsilon)l_c(\alpha\varepsilon)} \,. \tag{9}$$

Strain ε is represented by entry and exit thickness.

$$\varepsilon = \ln\left(\frac{H}{h}\right), \tag{10}$$

Under the strain ε and $\alpha\varepsilon$, the ratio of rolling reduction, resistance to deformation and contact arc length can be expressed as:

$$\frac{\Delta h(\varepsilon)}{\Delta h(\alpha\varepsilon)} = e^{\alpha\varepsilon - \varepsilon} \frac{He^\varepsilon - H}{He^{\alpha\varepsilon} - H} = e^{(\alpha-1)\varepsilon} \frac{e^\varepsilon - 1}{e^{\alpha\varepsilon} - 1} \,, \tag{11}$$

$$\frac{\sigma(\varepsilon)}{\sigma(\alpha\varepsilon)} = \frac{1}{\alpha^{a_3+a_4}}\left(e^{(\alpha-1)\varepsilon} \frac{e^\varepsilon - 1}{e^{\alpha\varepsilon} - 1}\right)^{-\frac{a_4}{2}} \,, \tag{12}$$

$$\frac{L_C(\varepsilon)}{L_C(\alpha\varepsilon)} = \sqrt{\frac{H - \dfrac{H}{e^\varepsilon}}{H - \dfrac{H}{e^{\alpha\varepsilon}}}} = \left(e^{(\alpha-1)\varepsilon} \frac{e^\varepsilon - 1}{e^{n\varepsilon} - 1}\right)^{1/2} \,. \tag{13}$$

Substituting the Equations (12) and (13) into Equation (9) and we get:

$$\frac{P(\varepsilon)}{P(\alpha\varepsilon)} = \frac{1}{\alpha^{a_3+a_4}}\left(e^{(\alpha-1)\varepsilon} \frac{e^\varepsilon - 1}{e^{\alpha\varepsilon} - 1}\right)^{\frac{1-a_4}{2}} \,. \tag{14}$$

In the place of strain $\alpha\varepsilon$, the rolling pressure is:

$$P(\alpha\varepsilon) = P(\varepsilon)\alpha^{a3+a4} / \left(e^{(\alpha-1)\varepsilon}\frac{e^{\varepsilon}-1}{e^{\alpha\varepsilon}-1}\right)^{\frac{1-a_4}{2}}. \quad (15)$$

It follows from Equation (15) that when α takes different values, we can get $P(\alpha\varepsilon)$ by $P(\varepsilon)$ which means that we have got other key points ($h(\alpha\varepsilon)$, $P(\alpha\varepsilon)$) needed by fitting curve, that is:

$$H / e^{\alpha\varepsilon}, P(\varepsilon)\alpha^{a3+a4} / \left(e^{(\alpha-1)\varepsilon}\frac{e^{\varepsilon}-1}{e^{\alpha\varepsilon}-1}\right)^{\frac{1-a_4}{2}}.$$

In order to get the third key point needed by conic fitting, this paper adopts $\alpha = 0.5$, then gets ($h(0.5\varepsilon)$, $P(0.5\varepsilon)$). Fitting the plastic curve of three points as follows: (H, *0*), ($h(0.5\varepsilon)$, $P(0.5\varepsilon)$), (h_{act}, P_{act}).

Suppose the conic shape is shown as follow:

$$y = b_0 + b_1 x + b_2 x^2. \quad (16)$$

Substitute (H, 0), (h_{act}, P_{act}) and ($h(0.5\varepsilon)$, $P(0.5\varepsilon)$) to the above formula, then we can get b_0, b_1 and b_2, as shown below:

$$b_2 = \frac{P_{act}(h(0.5\varepsilon)-H) - P(0.5\varepsilon)(h_{act}-H)}{(h(0.5\varepsilon)-H)(h_{act}-H)(h_{act}-h(0.5\varepsilon))}$$

$$b_1 = \frac{P_{act}}{h_{act}-H} - b_2(h_{act}+H) \quad . \quad (17)$$

$$b_0 = -b_1 H - b_2 H^2$$

The plastic coefficient is the tangency of the conic in actual rolling pressure point (h_{act}, P_{act}), that is:

$$Q = y'(h_{act}) = b_1 + 2b_2 h_{act}. \quad (18)$$

In Equation (17), the coefficients b_0, b_1 and b_2 of plasticity coefficient calculation need only to be related to plate thickness, rolling force, strain, and the deformation resistance coefficient a_3, a_4. The equations are simple in structure, but also the derived model based on the rolling mechanism is consistent with the definition of plasticity coefficient, without loss of computational accuracy.

The given above equation of plastic coefficient is simple and stable, it is easy to embed in the basic automatic system, and can achieve the fast and high-precision plastic coefficient calculation. Automation system for thickness control diagram is shown in Figure 2. Plastic coefficient on-line calculation module embedded before thickness

control algorithm. When plate exiting thickness is obtained based on the roll gap and the rolling force, compared with set thickness to obtain thickness deviation. Then the plastic coefficient and mill stiffness are sent to the AGC control model to obtain the compensation value of the hydraulic system, and achieve high-precision thickness control.
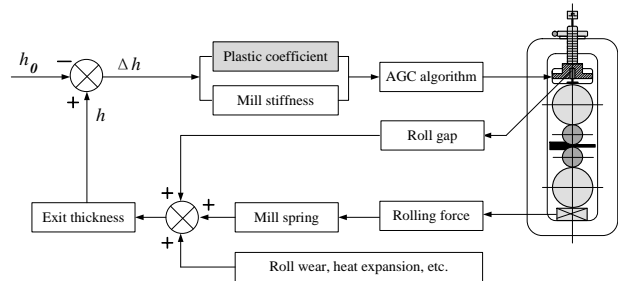
FIGURE 2 Schematic diagram of system for thickness control

3 Calculation process

The rolling force is constantly changing because of the fluctuations of the plate width, the rolling temperature and other factors during the rolling process. So the mill spring is constantly changing meanwhile. For this reason, the exit thickness of rolled pieces deviates from the set value. After the rolling mill gauge control model is put, the thickness correction is calculated based on the real-time measurement of the thickness changes, and then the compensation of the thickness fluctuation is implemented by adjusting the roll gap. The typical tendency of rolling force and roll gap is shown in Figure 3 in the rolling process.
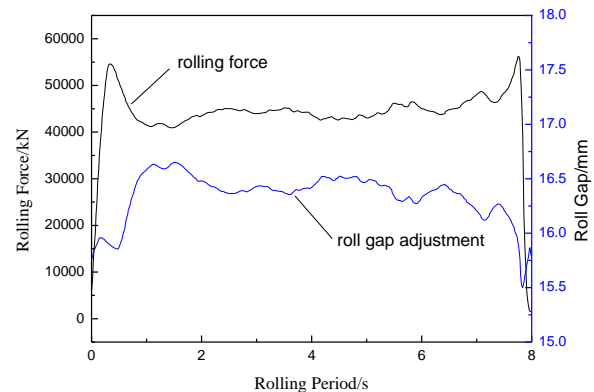

FIGURE 3 The rolling force and roll gap change tendency

To achieve accurate compensation for thickness fluctuations of rolled pieces, actual plastic coefficient must be calculated on the basis of actual rolling force and actual exit thickness. That is, the plastic coefficient of on-line calculation can maximally meet the requirements to eliminate thickness deviation.

Based on the above deduction of plastic coefficient calculation formula, we can get the following plastic coefficient's computational processes.

Step 1. Select deformation resistance model and force calculation equation. The calculation method of plastic

coefficient is derived based on the deformation resistance Equation (5) in the paper.

Step 2. Confirm the regression coefficients of deformation resistance corresponding to different steel grades by primary data input.

Step 3. Get the current pass rolled pieces' entry thickness. If the current rolling pass is the first pass, the entry thickness is slab thickness; otherwise the entry thickness is the exiting thickness of preceding pass.

Step 4. Making use of displacement and pressure sensor installed on the mill, measure the gap changes and rolling pressure in the rolling process and calculate rolled pieces' actual exit thickness.

Step 5. Make sure two known key points (H, 0) and (h_{act}, P_{act}) in the plastic curve, and deduce the third key point based on the rolling force formula.

Step 6. Bring the three known point coordinates to the quadratic curve equation, and use the conic to fit plastic curve and solve the equation's coefficients b_0, b_1 and b_2.

Step 7. In the fitting curve we have calculated the actual rolling pressure point (the actual exist thickness and actual rolling force), that is the tangency at (h_{act}, P_{act}) and then we can get the plastic coefficient.

Step 8. Put the actual plastic coefficient into thickness control model, calculate the compensation value of the roll gap, and execute this value by the control system to achieve the thickness compensation.

The calculation process diagram of plastic coefficient is shown in Figure 4.
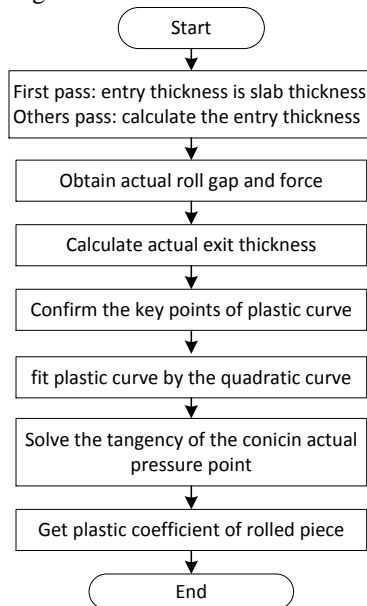
FIGURE 4 Calculation diagram of plastic coefficient

4 Result analysis

Use the above formula to calculate the plastic coefficient with Q345 and 25 mm thickness plate during the last rolling pass, and verify the accuracy of the plastic coefficient calculation formula. The details of rolling parameters are shown in Table 1.

TABLE 1 Rolling process parameters

Category	Value
Roller length	3000 mm
Work roll diameter	950 mm
Steel grade	Q345
Slab thickness	230 mm
Slab width	1550 mm
Slab length	2750 mm
Entry thickness	27.1 mm
Exit thickness	25.0 mm
Rolled piece width	2200 mm
Average temperature	915 ℃
Average speed	60 rpm
Average rolling force	11856 kN
Zero point force	20000 kN
Mill stiffness	7200 kN/mm

The on-line calculation result is shown as Figure 5 during the last rolling pass. From the figure we can see plastic coefficient changes constantly in the rolling process. The calculated plastic coefficient is less than the average value, so the thickness accuracy will be reduced by using average plastic coefficient in control system. Since the average plastic coefficient is calculated by process computer, the plate prediction thickness is poor, and the difference between real plastic coefficient and average plastic coefficient is bigger.
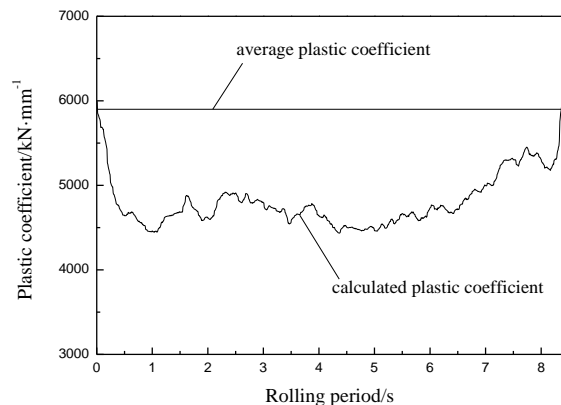
FIGURE 5 Schematic diagram of plastic coefficient change

The variation of calculated plastic coefficient shows the fluctuation of plate thickness, width, and temperature in the rolling process, the fluctuation results illustrate the change in plate plasticity, and reflected by plastic coefficient. The plate's head and tail temperature is lower than middle temperature, therefore the head and tail's plastic coefficient is bigger than middle position. The calculation result also reflects this changing process.

5 Conclusions

1) The plastic coefficient on-line calculation model is derived based on the rolling mechanism model, and the calculation formula is not influenced by the production site's complex factors. The result is stable and is easy to embed in the automatic system of hot rolling. The calculation results of plastic coefficient can be directly transferred to the thickness control algorithm and applied to high-precision AGC control system.

2) Based on determining the key points of plastic curve, the method that using quadratic curve instead of plastic curve is developed in the paper, and then the plastic coefficient is solved. The calculation formula of plastic coefficient is simple and fast in calculation speed. If more precision plastic coefficient is needed, the above formula can be adopted to determine the more critical point on the plastic curve, and use high-order curve fitting plastic curve.

3) The plastic coefficient got by the paper's calculation is the tangency of the actual rolling pressure point in the plastic curve. It exactly matches the plastic coefficient's definition, and the computational accuracy is higher than traditional method. Compared to the traditional method to solve plastic coefficient mean, its solving accuracy is increased by more than 6%, and the coefficient is more suitable for high-precision thickness control system.

References

[1] Gao Y J, Wang Y Q, Kong X D 1998 The study on dynamic model of hydraulic AGC system in a strip mill *Journal of Yanshan University* **22** 259-62 *(in Chinese)*

[2] Wang Z L, Tong C N, Sun Y K 2005 Real-time simulator of AGC systems for hot-rolling mill, *Journal of University of Science and Technology Beijing* **27** 600-603 *(in Chinese)*

[3] Zhong Y F, Tan S B, Xu X H 2009 Research and application of the feedforward-AGC system in hot strip rolling mills *Journal of Northeastern University (Natural Science)* **30** 169-71 *(in Chinese)*

[4] Wang J, Li J P, Wang G D 2002 Mill modulus control on pressure closed loop *Control Decisions* **17** 126-8 *(in Chinese)*

[5] Zhang J Z 1989 New method for measuring plastic coefficient of rolling stock and its application, *Iron & Steel* **24** 33-7 *(in Chinese)*

[6] Jiang L Y, Zhao C J, Zhang J Z 2011 Methods for calculating plastic coefficient of rolling stock and their errors *Mechanical Engineering & Automation* **5** 102-4 *(in Chinese)*

[7] Zhang D H, Zhang W X, Yan D 2010 Study and application of q prediction feedforward agc in hot tandem rolling *Iron & Steel* **45** 51-4 *(in Chinese)*

[8] Linghu K Z, He A R, Yang Q 2007 Feedback-decoupling for a combined shape and gauge control system in hot strip rolling, *Journal of University of Science and Technology Beijing* **29** 338-41 *(in Chinese)*

[9] Takashi O, Naoki S, Toshiki Y 1995 Adaptive technology for thickness control of finisher set-up on hot strip mill *ISIJ International* **35**(1) 42-9

[10] Prodip B 1996 Modeling and simulation of hydraulic gap control system in a hot strip mill *ISIJ International* **36**(5) 553-562

[11] Zhang D H, Li X, Zhang H, Han J Z 2008 Algorithmic design and application of pure lag compensation controller for monitor AGC via gap *Iron & Steel* **43**(6) 52-5 *(in Chinese)*

[12] Hu X L, Zhao Z, Liu X H, et al 2007 Analysis of oversize in thickness of plate at head and tail ends, *Iron & Steel* **42** 50-2 *(in Chinese)*

[13] Li H J, Xu J Z, Wang G D 2009 High-precision rolling force prediction model for hot strip continuous rolling process *Journal of Northeastern University (Natural Science)* **30** 669-72 *(in Chinese)*

Authors

Chunyu He, born in March, 1973, Shenyang, Liaoning, China

Current position, grades: Northeastern University, associate professor.
University studies: PhD degree in materials processing engineering, Northeastern University, Shenyang, China, 2009.
Scientific interests: rolling theory, optimization of rolling technology, adaptive control.
Publications: 15.

Zhijie Jiao, born in July, 1976, Shenyang, Liaoning, China

Current position, grades: Northeastern University, associate professor.
University studiesi: PhD degree in materials processing engineering, Northeastern University, Shenyang, China, 2006.
Scientific interests: rolling theory, development of rolling model.
Publications: 20.

Di Wu, born in June, 1952, Shenyang, Liaoning, China

Current position, grades: Northeastern University, professor.
University studies: PhD degree in materials processing engineering, Northeastern University, Shenyang, China, 1987.
Scientific interests: rolling theory and technology of section steel, fast cooling technology.
Publications: 60.

Error modelling of depth estimation based on simplified stereo vision for mobile robots

Bo Jin*, Lijun Zhao, Shiqiang Zhu

The State Key Lab of Fluid Power Transmission and Control, Zhejiang University, Hangzhou, 310027, China

Abstract

Depth estimation is the precondition in obstacle avoidance for mobile robots. To improve the obstacle detecting effectiveness and quickness in poor-textured backgrounds, we used the centroid abscissa difference of corresponding obstacle region in image pairs as parallax to estimate obstacle depth. The error of parallax and depth were studied analytically and numerically. Wood blocks of different shapes and sizes were used for demonstrating the relationship between estimated depth and actual depth. A quadratic function model was obtained after experiments. Although the depth estimation error was relatively higher compared to conventional grayscale correlation-based method, the proposed method was expected to satisfy the accuracy requirement of depth estimation for common mobile robots.

Keywords: error model, depth estimation, stereo vision, radial distortion

1 Introduction

Depth estimation is prerequisite to obstacle avoidance for mobile robots and various sensors can be adopted [1, 2]. Being lightweight, power-efficient and inexpensive, stereo cameras are preferred in range finding [3, 4]. However, cameras need accurate calibration and algorithms rely seriously on scene textures. To guarantee estimation accuracy, image rectification is inevitable. Meanwhile, stereo matching is needed which uses similarity functions including region descriptors [5, 6] and feature descriptors [7, 8]. These procedures lead to heavy calculation burden. Moreover, high mismatching rate happen to applications in poor-textured scenes. In this paper, we presented a strategy for depth estimation in poor-textured scenes. Conventional distortion rectification was avoided and density-based matching was substituted with calculating the centroid abscissa difference of corresponding obstacle regions in image pairs. That difference was taken as parallax to estimate obstacle depth.

Zou analysed the influence on depth estimation by hardware system error, camera calibration error, feature extraction error and stereo matching error [9]. Rodriguez analysed the quantization error of stereo vision system and presented the probability density functions [10]. Fooladgra gave a geometrical approach to estimate the amount of localization error [11]. However, the studies mainly focused on theoretical analysis. Some models help to minimize depth estimation error but increase computational burden, so they are more suitable for preliminary theoretical design rather than real-time application. Rodriguez indicated that acceptable error should be decided before selecting relevant parameters [10]. Llorca suggested that a trade-off should be reached

between accurate estimation and other parameters [12]. So we suggested estimation strategy be based on acceptable error in specific applications. For mobile robots, less depth estimation accuracy is acceptable in return for real-time performance. The lost accuracy can be balanced by setting safe margin for obstacle areas referencing the strategies of UAV obstacle avoidance and route planning which represents the obstacle areas as circles [13]. As to the errors of parallax and depth estimation in our strategy, analytical and numerical studies were given in Sections 2 and 3.

2 Depth estimation error analysis

The strategy we proposed for obstacle depth estimation included three steps. First, image pairs were block-segmented in space domain. Then feature vector of candidate obstacle areas were extracted. Finally, the centroid abscissa difference of corresponding obstacle areas was taken as parallax for depth estimation. There were several studies on image block segmentation [14]. Meanwhile, the components of feature vector are mainly coordinates easy to be extracted. So we mainly focused on modelling of obstacle depth estimation error, assuming that images are segmented correctly and the extracted centroids coincide with their actual position.

Considering point $P(x_w, y_w, z_w)$ in the world coordinate system (Figure 1), if the projective points of P in left and right image planes are $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$ in each image coordinate system, then the theoretical depth of P can be expressed based on triangulation method:

---

* *Corresponding author* e-mail: bjin@zju.edu.cn

$$z = \frac{B \cdot f}{\lambda \cdot d} , \tag{1}$$

where B is baseline length and f is camera lens focal length. The parallax of corresponding points is denoted as d. The size of pixel is λ. Assuming $x_l > x_r$, then we have:
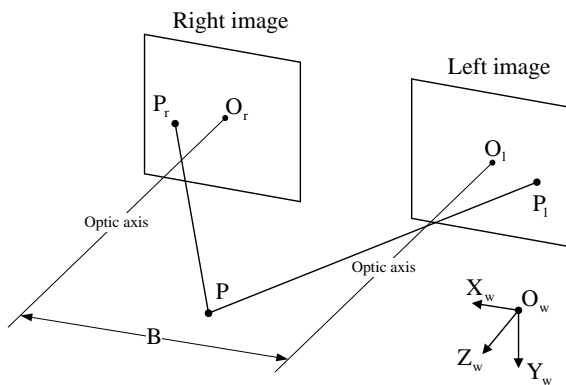
$$d = x_l - x_r . \tag{2}$$

FIGURE 1 The stereo vision system model

The pixel size is given in camera manual. The focal length can be calculated by calibration. The measurement accuracy of B is easy to be guaranteed. So we take only the error of parallax d into account while analysing the error of depth z.

The error of parallax caused by lens distortion is denoted as Δd, then the relationship between measured parallax d' and ideal parallax d is given by:

$$d' = d + \Delta d . \tag{3}$$

Thus, the estimated depth can be calculated as:

$$z' = \frac{B \cdot f}{\lambda \cdot d'} . \tag{4}$$

Using Equations (1) and (3), the error of depth estimation can be written as:

$$\Delta z = z - z' = \frac{B \cdot f}{\lambda \cdot d} - \frac{B \cdot f}{\lambda \cdot d'} =$$
$$\frac{B \cdot f}{\lambda \cdot d} - \frac{B \cdot f}{\lambda \cdot (d + \Delta d)} = z^2 \frac{\lambda}{B \cdot f} \cdot \frac{\Delta d}{1 + \Delta d / d} . \tag{5}$$

Based on Equation (5), we expected that within certain range $(z_{\min}, z_{\max})$ for z, if:

$$g(\Delta d, d) = \frac{\lambda}{B \cdot f} \cdot \frac{\Delta d}{1 + \Delta d / d} \approx C_e \text{(constant)} ,$$

the following relation will become true:

$$\Delta z = C_e \cdot z^2 . \tag{6}$$

Among common errors [9], calibration error and matching error are not involved in this paper. Ignoring the hardware system error and feature extraction error, we analyse the parallax error caused by camera lens distortion.

Camera lens distortions mainly include radial distortion, decentring distortion and prism distortion [15]. We considered only the radial distortion here because it dominates in all the distortions.

In image coordinate systems, the projective points of optical centre for left and right camera lens are denoted as $O_l(x_l, y_l)$ and $O_r(x_r, y_r)$, and we have the following commonly used equations:

$$x_l' - u_l = (x_l - u_l) +$$
$$k_l(x_l - u_l)[(x_l - u_l)^2 + (y_l - v_l)^2] , \tag{7}$$

$$x_r' - u_r = (x_r - u_r) +$$
$$k_r(x_r - u_r)[(x_r - u_r)^2 + (y_r - v_r)^2] , \tag{8}$$

where x_l' and x_r' are the actual abscissas in left and right images containing distortions, k_l and k_r are the first order distortion coefficients of the left and right camera lenses.

Using Equations (2), (3), (7) and (8), we obtained:

$$\Delta d = (x_l' - x_r') - (x_l - x_r) =$$
$$k_l(x_l - u_l)[(x_l - u_l)^2 + (y_l - v_l)^2] -$$
$$k_r(x_r - u_r)[x_r - u_r]^2 + (y_r - v_r)^2] . \tag{9}$$

We assumed that the lens parameters of the left and right cameras are similar and neglect the centroid ordinate difference of corresponding blocks in the left and right images:

$$u_l \approx u_r = u , \ v_l \approx v_r = v , \ y_l \approx y_r = y , \ v_l \approx v_r = v . \tag{10}$$

Combining Equations (2) and (10), Equation (9) can be written as follows:

$$\Delta d = k(x_l - u)[(x_l - u)^2 + (y - v)^2] -$$
$$k(x_r - u)[(x_r - u)^2 + (y - v)^2] =$$
$$k(d^3 + 3d^2 - d(x_r - u)^2 + d(y - v)^2) . \tag{11}$$

Finally, we have:

$$g(\Delta d, d) = g(d, x_r, y) \frac{\lambda}{B \cdot f} \cdot \frac{\Delta d}{1 + \Delta d / d} =$$
$$\frac{\lambda}{B \cdot f} \cdot \frac{k(d^3 + 3d^2 - d(x_r - u)^2 + d(y - v)^2)}{1 + k(d^2 + 3d - (x_r - u)^2 + (y - v)^2)} . \tag{12}$$

a) $k = 0.0001$

b) $k = 0.0005$

c) $k = 0.0025$

d) $k = 0.0125$

e) $k = 0.0625$

f) $k = 0.3125$
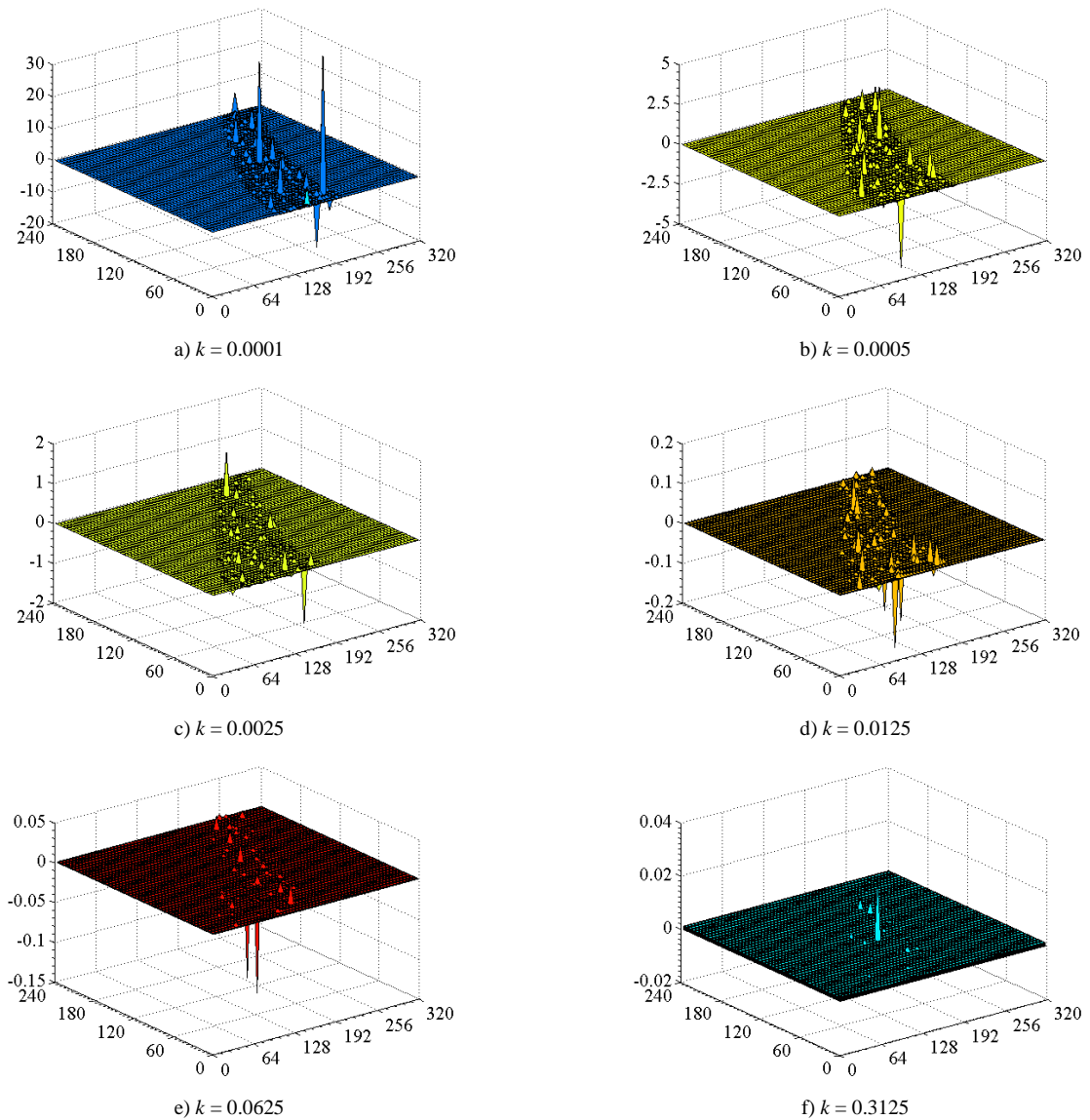
FIGURE 2 Value range of $f(d, x_r, y)$

3 Numerical simulation and experiment

To observe the range of $g(d, x_r, y)$, image resolution was set as 640×480. Denoting $x_{dist} = x_r - u$, $y_{dist} = y_r - v$, then $x_{dist} \in (1, 319)$, $y_{dist} \in (1, 239)$. The value of k which means the first order radial distortion coefficient is less than 1, so we took a candidate value set as {0.0005, 0.0025, 0.0125, 0.0625}. The camera system we adopted was STH-DCSG-VAR from Videre Design, parameters were obtained as $f = 4.6\,mm$, $\lambda = 0.006\,mm$, $B = 135\,mm$.

By setting the depth estimation extent as 600mm to 4000mm, the range of d was determined as (25.875, 172.5) based on Equation (1). Then the simulation results

were obtained (Figure 2).

Points in every image from Figure 2 represented the overlap of data when d varied from 1 to 319. It can be learned that the values of $g(d, x_r, y)$ kept approximately invariable (flat areas) and the change of k only exerts significant influence on the value of breaking points which appeared sparsely. The depth estimation value surging caused by the breaking points can be suppressed by using Kalman Filter. Therefore, Equation (6) is basically suitable for depth estimation error modelling in engineering applications.

Regular blocks of different shapes and sizes were used as objects to evaluate the proposed error model for depth estimation. Part of the results is given in Table. 1.

TABLE 1 Data from testing experiment (Unit: cm)

Cube				Sphere			
Side length=7		Side length=10		Radius=3.5		Radius=5	
z	z'	z	z'	z	z'	z	z'
530	486	550	416	730	742	545	516
930	929	950	856	1130	1205	945	960
1330	1406	1350	1327	1530	1700	1345	1432
1730	1923	1750	1821	1930	2228	1745	1943
2130	2451	2150	2368	2330	2799	2145	2464
2530	3048	2550	2963	2730	3416	2545	3060
2930	3690	2950	3589	3130	4142	2945	3652
3330	4378	3350	4249	3530	4768	3345	4306

All the testing data were utilized for regression. The quadratic regression curve was shown in Figure 3. The regressed coefficients for model $z' = C_1 z + C_2 z^2$ were $C_1 = 0.997$ $C_1 = 6.8 \times 10^{-5}$. It can be seen that the regression result basically coincided with the proposed error model.
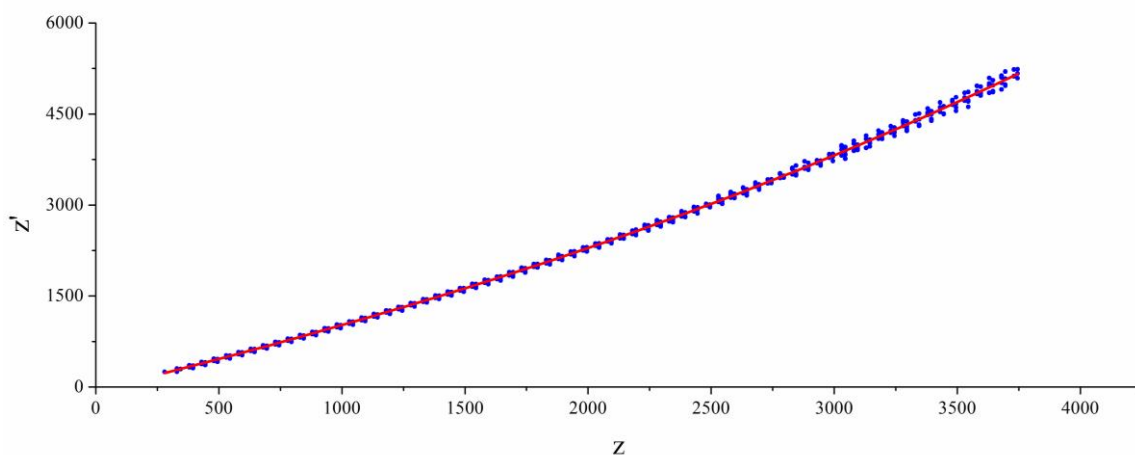
FIGURE 3 All data of validation experiment

4 Conclusions

To facilitate the stereo vision-based obstacle avoidance for mobile robot in poor-textured scenes, we proposed a strategy which needed not the conventional distortion rectification and area-based stereo matching. By image segmentation and obstacle centroid coordinates extraction, abscissa difference was used as parallax to estimate the depth of obstacles. Analytical deduction and numerical simulation were given to prove the feasibility of the hypothesis model. Regression validated the error model in subsequent experiment.

The depth estimation error was larger than the conventional method but still acceptable for mobile robot obstacle avoidance while the processing and calculating time was obviously shortened. The proposed strategy was feasible in general. In future works, the authors will use the proposed strategy combining Kalman Filter for mobile robot indoor obstacle avoidance. Efforts will also be put into improving the proposed model to reduce depth estimation error and processing time.

Acknowledgements

References

[1] Lai X B, Wang HS, Xu Y H 2012 A Real-time Range Finding System with Binocular Stereo Vision *International Journal of Advanced Robotic Systems* **9**(27) 1-9

[2] Wang T, Zheng N N, Xin J M, Ma Z 2011 Integrating Millimeter Wave Radar with a Monocular Vision Sensor for On-road Obstacle Detection Applications *Sensors* **11**(9) 8992-9008

[3] Stelzer A, Hirschmuller H, Gorner M 2012 Stereo-vision-based Navigation of a Six-legged Walking Robot in Unknown Rough Terrain *International Journal of Robotics Research* **31**(4) 381-402

[4] Antunes M, Barreto J P, Premebida C, Nunes U 2012 Can Stereo Vision Replace a Laser Rangefinder? *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* Vilamoura Portugal Oct 2012 5183-90

[5] Donate A, Liu X, Collins E G 2012 *IEEE Transactions on System, Man, and Cybernetics-Part B: Cybernetics* **41**(1) 183-95

[6] Verma R, Singh H S, Verma A K 2013 Depth Estimation from Stereo Images Based on Adaptive Weight and Segmentation *Journal of The Institution of Engineers* **93**(4) 223-9

[7] Fickel G P, Jung C R, Malzbender T, Samadani R, Culbertson B 2013 *IEEE Transaction on Image Processing* **22**(9) 3353-65

[8] Brown M, Hua G, Winder S 2011 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(1) 2011 43-57

[9] Zou XJ, Zou HX, Lu J 2012 Virtual Manipulator-based Binocular Stereo Vision Positioning System and Errors Modelling *Machine Vision and Applications* 23(1) 43-63
[10] Rodriguez J J, Aggarwal J K 1990 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(5) 467-70
[11] Fooladgar F F, Samavi S, Soroushmehr S M R, Soroushmehr R 2013 *IEEE Sensors Journal* **13**(11) 4236-46
[12] Llorca D F, Sotelo M A, Parra I 2010 Error Analysis in A Stereo Vision-Based Pedestrian Detection Sensor for Collision Avoidance Applications *Sensors* **10**(4) 3741-58
[13] Chen H D, Chang K C, Agate C S 2013 *IEEE Transactions on Aerospace and Electronic Systems* **49**(2) 840-56
[14] Xiong G M, Li X, Xi J Q 2010 Object Distance Estimation Based on Stereo Vision and Color Segmentation with Region Matching *Advances in Visual Computing* **6455** 368-76
[15] Wang J H, Shi F H, Zhang J 2008 A New Calibration Model of Camera Lens Distortion *Pattern Recognition* 41(2) 607-15

Authors

Bo Jin

Current position, grades: Doctor of Mechanical Engineering, Professor, Zhejiang University.
University studies: B. E. and Ph. D on Fluid Power Transmission and Control awarded by Zhejiang University respectively in 1993 and 1998.
Research interest: deep-sea mechatronic equipment, electric hydraulic control system, intelligent robot, embedded system and applications.
Publications: Over 70 articles, 16 patents and software copyrights, 1 project awarded National Technical invention second prize and 2 projects awarded Provincial S&T progress first prize (Principle person)

Lijun Zhao

Current position, grades: Ph. D. candidate on Mechatronic Engineering, Zhejiang University.
University studies: B. E. (Mechanical Engineering), Lanzhou Jiaotong University (2003). M.E. (Agriculture Mechanization Engineering), Jilin University (2008).
Research Interests: Machine vision and application, mobile robot navigation and control, industrial automation and nondestructive detection.

Shiqiang Zhu

Current position, grades: Doctor of Mechanical Engineering, Professor, Vice-Director of Zhejiang University Robotics Research Center, Vice-President of Zhejiang Ocean University.
University studies: B. E., Fluid Power Transmission and Control, 1988, Zhejiang University; M. E., Mechatronic Engineering, 1991, Beijing Institute of Technology; Ph. D., Mechatronic Engineering, 1994, Zhejiang University.
Research interest: robotics, electro-hydraulic automatic control system, semi-physical simulation technology.
Publications: over 80 articles, 14 patents.

A study of characteristics extraction of dynamic pressure signals in pipeline based on EMD

Wei Liu[1, 2*], Hongzhao Liu[1]

[1]*School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an, Shaanxi, 710048, China*

[2]*Department of Computer Science and Technology, Shaanxi Xueqian Normal University, Xi'an, Shaanxi 710100, China*

Abstract

Characteristic selection is a key to accurate signal recognition. Specific to the signal recognition of dynamic pressure waves in pipelines, this article proposes to use EMD method to decompose dynamic pressure signals into a series of IMFs, adopt the correlation theories of signals to determine and eliminate noise, retain valid IMFs from 4th to 8th level and extract kurtosis, energy and statistical characteristics. The experiment analyses and verifies that the energy or statistical characteristics (mean variance, mean standard deviation, mean range and mean quartile range) can effectively represent signals.

1 Introduction

The research of pipeline leakage detection technology started from 1970s. Due to the differences in measuring media, there are mainly two detection methods: direct detection and indirect detection [1].The negative pressure wave method used in indirect detection method only requires pressure transmitters on both sides of the pipeline to detect the real-time changes in pressure. It is widely used in China as it does not need a mathematic model, is convenient in construction and maintenance, costs relatively low and fits China's pipe network [2]. When a pipeline leaks, the pressure in the leaking point decreases out of a sudden and surrounding fluid will flow towards it. Such a phenomenon is called the wave fluctuations of negative pressure wave which spreads to both sides of the pipeline when leakage takes place. It has characteristics of long distance of spread, instant responsiveness and reliable signals. Actual pressure signals, including normal pressure wave, negative pressure wave and other effective signals and noises generated by environment, are collected by pressure transmitter. They are non-linear, non-stationary, vibrating signals with noise and can be analysed and recognized after de-noising.

Empirical Mode Decomposition (EMD), a non-linear, non-stationary signal processing method, was proposed in 1998 by Dr. Nomen E, Huang (Huang E) and his colleagues in Goddard Space Flight Centre in National Aeronautics and Space Administration (NASA) [3]. It has been applied in damage detection, biotechnology, filtration, de-noising and other areas. EMD aims to decompose adaptively a time series signal, according to its own time measurement, into an Intrinsic Mode Function (IMF) with both orthogonally and completeness [4]. These IMFs are decomposed according to the frequency sequence from high to low. Analysis of all variants is useful for measuring the signals' characteristics more effectively and accurately. Such an assembly of IMFs according to the frequency equals to the high-pass, low-pass and band-pass filtration of all signals. Among common filtration methods, Fourier transform primary function is a trigonometric function , suitable for time-frequency analysis of stationary signals. The short-time Fourier transform (STFT), Gabor Transform, wavelet analysis and other methods were proposed to analyse non-stationary signals. However, their primary functions are fixed without adaptively, and the de-noising effects of wavelet analysis rely on the artificial selection of wavelet basis and decomposition layers [5]. Therefore, EMD method with adaptively is more suitable for the processing and analysis of the instant pressure signals in non-linear and non-stationary pipelines.

This article adopts EMD method to decompose the pressure signals in petroleum pipeline into a series of IMFs and measure the signal fluctuations of the IMFs under different frequencies. It analyses the validity of IMFs according to autocorrelation and cross-correlation with original signals and composes all valid IMFs to constitute de-noising signals. It proposes to pick up kurtosis [6], energy and statistical characteristics of the de-noising signals and, through SVM method, compares the effectiveness of these characteristics to describe pressure signals.

---

[*] *Corresponding author* e-mail: yoyotianxia@163.com

455

2 De-noising of dynamic pressure signals based on EMD

The pipeline pressures signals that dynamic pressure transmitters collect are non-linear, non-stationary, noise-including, vibrating signals. With EMD, signals can be decomposed into a series of IMFs according to the frequency from high to low. Choose IMFs within valid frequencies according to signals characteristics to compose new signals. Such a process equals to the filtration of original signals.

2.1 DECOMPOSITION OF DYNAMIC PRESSURE SIGNALS BASED ON EMD

The EMD method assumes that [3] all signals composed of a series of different IMFs. Each IMF can be linear or non-linear, stationary or non-stationary. Each IMF should satisfy two requirements: the number of extreme points and the number of zero points of the curves equal or have a maximum difference of 1; on any point in the curve, the average of the maximum extreme point and minimum extreme point of the envelop equals zero, namely that the envelop is symmetrical on the time axis. EMD method decomposes signals according to the time characteristics of the data series. There is no need to present a primary function. It is a data-driven signal analysis method with adaptively, suitable to process frequency-transient signals, non-stationary and non-linear signals.

When the negative pressure wave method detects pipeline leakages, the pressure wave signals that the pressure transmitters at both sides of the pipeline collect are non-stationary. The EMD method, when decomposing signals, can analyse IMFs with different frequencies and different vibrating modes, therefore requiring wave information within the signals. After decomposition, original signal $x(t)$ is decomposed into a combination of finite IMFs and one residual $r(t)$.

$$x(t) = \sum_{i=1}^{n} imf_i(t) + r(t). \tag{1}$$

Each IMFs is decomposed according to the frequency from high to low and represents the fluctuations of signals under different frequencies. The residual $r(t)$ represents the tendency of the signals.

2.2 SIGNAL ANALYSIS AND DE-NOISING BASED ON EMD

Recently, EMD method has been used in pipeline leakage detection technology to reduce noise interference and to analyse valid signals. As EMD uses cubic spline interpolation data structure to compose the upper and lower envelops of IMFs and due to the requirement of symmetry on the time axis during the decomposition, there exist pseudo-IMFs among all the IMFs decomposed that will interfere the signal analysis. The 6th and 7th articles cited uses the former 6 and former 4 IMFs respectively to re-compose signals, reduce low-frequency signals and residual signals and maintain high-frequency IMFs. The 8th article calculated the correlation of all IMFs with original signals, and used three IMFs with highest correlation coefficients to de-compose negative pressure wave signals. The 5th article selected IMFs from 5th to 8th levels and residual signals to compose valid signals. It used the cross-correlation and auto-correlation of signals to measure the valid IMFs within all the actual signals in the test oil fields.

The 9th article proposed the detection of noise by analysing the cross-correlation and auto-correlation of IMFs and original signals. The correlation coefficient of white noise and original signals is zero. If the correlation coefficient $R_{x\,imf_i}(\tau)$ of a certain $imf_i(t)$ and original signal $x(t)$ is smaller, then such a function can be noise:

$$R_{x\,imf_i}(\tau) = E\left[x(t)imf_i(t+\tau) \right], \tag{2}$$

where if the correlation coefficient $R_{\ddot{imf}_i(\tau)}$ of suspected noise $\dddot{imf}_i(t)$ is small at each point except the maximum at point zero, such a function can be determined to be noise. Therefore, it is verified that the correlation coefficient of white noise is zero at each point except the maximum at point zero.

$$R_{\dddot{imf}_i(\tau)} = E\left[\dddot{imf}_i(t)\dddot{imf}_i(t+\tau) \right]. \tag{3}$$

Due to the existence of decomposition error of regional wave, last few functions are always pseudo-functions. As pseudo-functions have low correlation with original signals, they can be excluded without influencing the analysis result.

2.3 CORRELATION ANALYSIS AND DE-NOISING OF FIELD-MEASURED SIGNALS

The characteristic of field-measured signals is that the intrinsic signals are overwhelmed by noise and that signals tend to perform like high-frequency noise. Therefore, correlate the decomposed IMFs with original signals. Those IMFs with high correlation coefficients represent noise and those with the lowest are usually pseudo-functions. If the correlation coefficient $R_{\ddot{imf}_i(\tau)}$ of IMFs is small except the maximum at point zero, it can be determined to be noise.
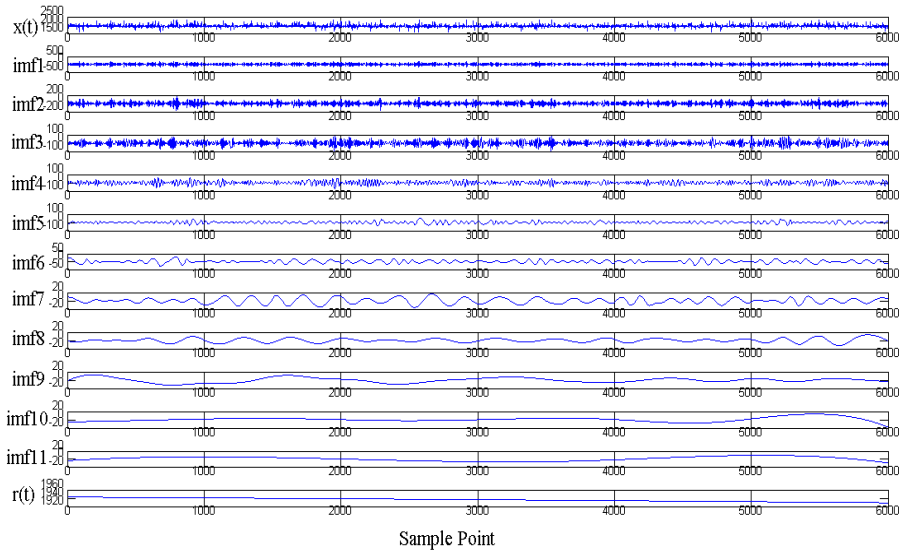
456

FIGURE 1 Normal Signals and its IMFs after EMD

Normal signals and its IMFs after EMD are shown in Figure 1, where $x(t)$ is the collected signal. IMF 1 to IMF 11 are IMFs after EMD at various levels. They are decomposed according to the frequency from high to low and the decomposed form is unique. Each IMF represents the changes of signal amplitude under a certain frequency with the changes of sampling time. $r(t)$ is the residual which represents the tendency of signals. The calculation of the cross-correlation coefficients of each IMF and original signals $x(t)$ is shown in Table 1. In the table, those IMFs with high correlation coefficients at the upper levels are noise; those with low correlation coefficients at the lower levels are pseudo-functions. Calculate and draw an auto-correlation graph of IMFs, as shown in Figure 2. It can be seen that, the IMFs decomposed earlier have more obvious characteristics of white noise.

TABLE 1 Cross-Correlation Coefficients of Normal Signals and its IMFs

$R_{0,1}$	$R_{0,2}$	$R_{0,3}$	$R_{0,4}$	$R_{0,5}$	$R_{0,6}$	$R_{0,7}$	$R_{0,8}$	$R_{0,9}$	$R_{0,10}$	$R_{0,11}$	$R_{0,12}$
0.6272	0.4285	0.2942	0.2072	0.1407	0.1106	0.0761	0.0448	0.0688	0.0759	0.0786	0.0165
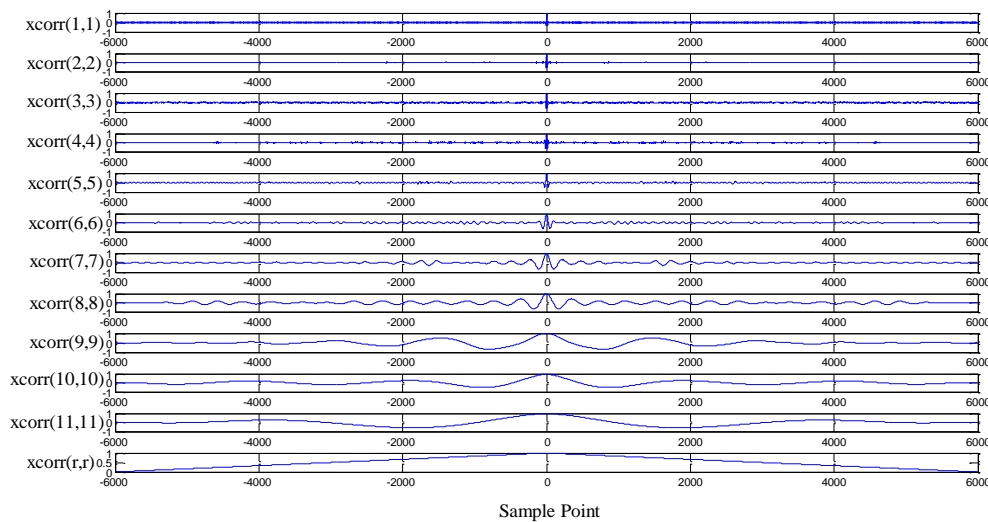
FIGURE 2 The Correlation of Normal Signals and its IMFs

The leakage signals and its IMFs after EMD are shown as Figure 3. The leakage signals are also decomposed into IMFs of 11 layers and one residual $r(t)$. The cross-correlation coefficients of each IMF and the original signal $x(t)$ are shown in Table 2. The auto-correlation of each IMF is shown in Figure 4.

Sample Point
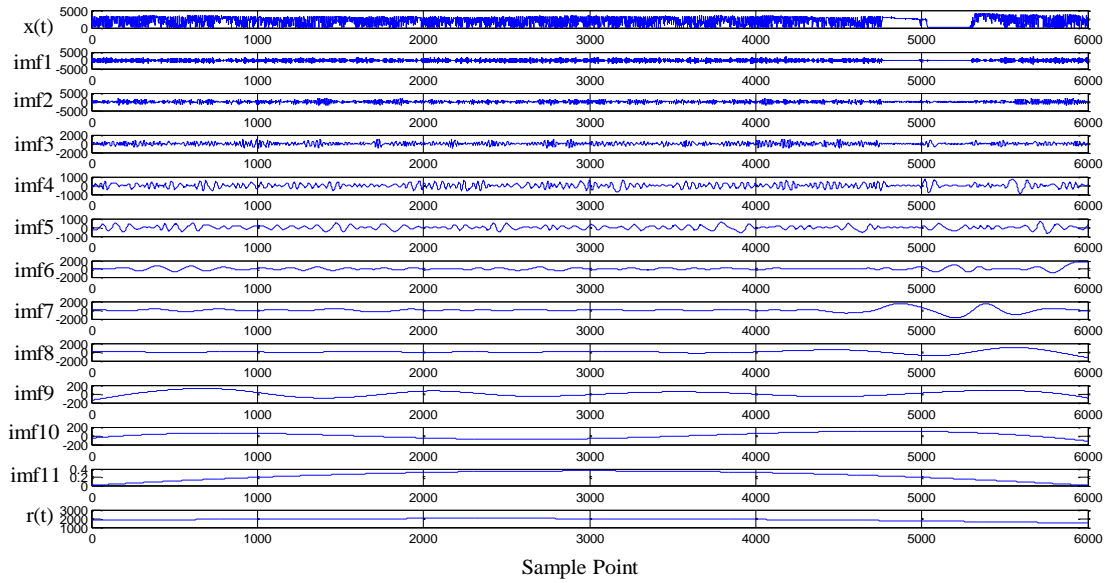
FIGURE 3 Leakage Signal and its IMFs after EMD

TABLE 2 The Cross-Correlation Coefficients of Leakage Signals and its IMFs

$R_{0,1}$	$R_{0,2}$	$R_{0,3}$	$R_{0,4}$	$R_{0,5}$	$R_{0,6}$	$R_{0,7}$	$R_{0,8}$	$R_{0,9}$	$R_{0,10}$	$R_{0,11}$	$R_{0,12}$
0.5793	0.4870	0.2226	0.1982	0.2121	0.1829	0.3549	0.1978	0.0141	-0.0024	0.0110	0.0223

Sample Point
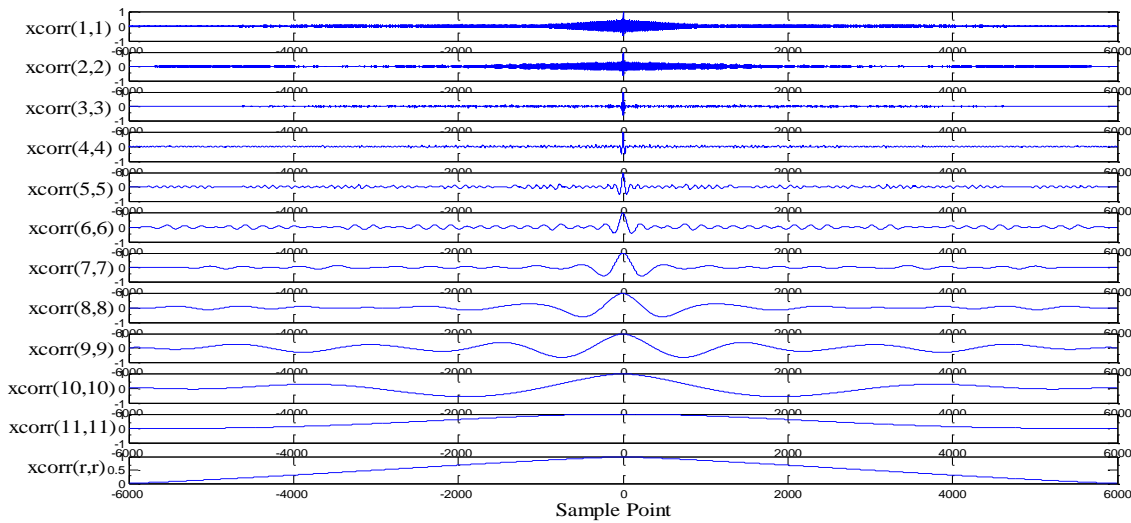
FIGURE 4 Cross-correlation of Leakage Signals and IMFs

To analyse comprehensively, eliminate the IMFs with high frequencies and white noise characteristics and pseudo-IMFs with low frequencies and low amplitude, retain IMFs from 4th to 8th level and reconstruct signals. The signals retain the characteristics of original signals after de-noising and adopting transform the average value of function to nought, as shown in Figures 5 and 6.
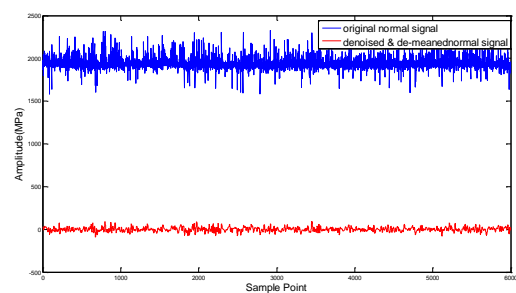
FIGURE 5 Normal Signal

458
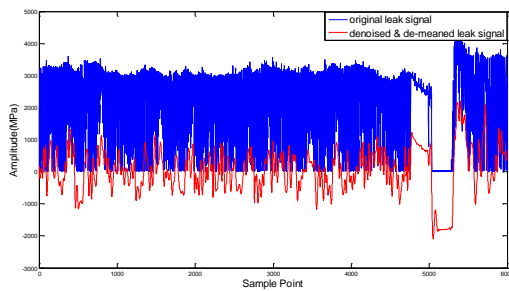
FIGURE 6 Leakage Signal

3 Three characteristic extraction methods

After de-noising and de-meaning, the turbulence characteristic of dynamic pressure wave signals is very obvious, with noticeable characteristics of a one-dimensional time series. The signal characteristics of normal signals and leakage signals are remarkably different. The following analysis selects signal characteristics from kurtosis, energy and statistics, and compares the effective characteristic description method of signals.

3.1 KURTOSIS CHARACTERISTIC

Kurtosis is a non-dimensional parameter to describe the peak of waves, often used in fault diagnosis. The 6th and 7th articles used the normalized kurtosis of IMFs as the main characteristic parameter of dynamic pressure signals. This article uses the normalized kurtosis of the IMFs from 4th to 8th levels of an oil field measured signals as the signal descriptive characteristic.

The Equation (4) is for kurtosis calculation. μ_{imf_i} and σ_{imf_i} are the mean value and the standard deviation of the IMF on i level. $E(t)$ calculates the expected value of the variant, $i \in [4,8]$.

$$K_i = \frac{E\left(imf_i - \mu_{imf_i}\right)^4}{\sigma_{imf_i}^{\ 4}} . \tag{4}$$
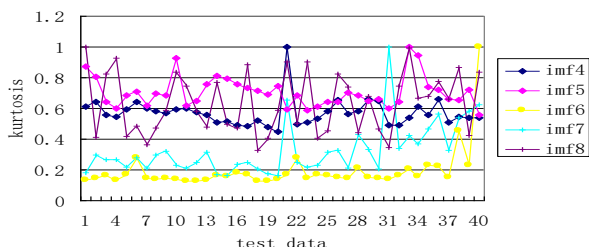
FIGURE 7 Kurtosis of IMF of Test Data

Figure 7 illustrates the zoomed kurtosis (between 0 and 1) of IMFs from 4th to 8th level from an EMD of 40 pairs of test data. The previous 30 pairs are normal data while the latter 10 are leakage data.

3.2 ENERGY CHARACTERISTIC

The 2nd article pointed out that Hilbert-Huang transform could provide accurately the conjoint analysis of signal energy with changes in frequency and time. Use the normalized energy of IMFs from 4th to 8th level as signal descriptive signal. The signal characteristics in Graph 7 are shown in Figure 8.
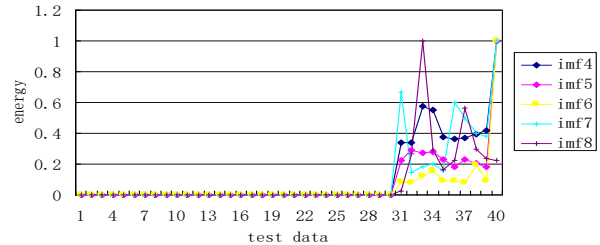
FIGURE 8 Energy Characteristics of IMF of Test Data

3.3 STATISTICAL CHARACTERISTIC

Statistical characteristic is an efficient method to describe time series data with even intervals. Use the mean value of statistical characteristics of IMFs from 4th to 8th level to describe the signals. The signal characteristics (Figure 7) are shown in Figure 9.
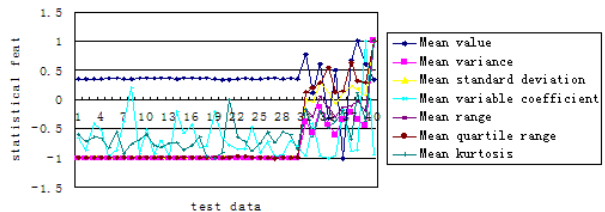
FIGURE 9 Average Statistical Characteristics of IMFs in Test Data

3.4 CHARACTERISTIC EXTRACTION

The 10th article proposed 3 standards to select leakage signals:

1) given that data length stays the same, wherever the leakage signals appears in this data, its characteristic values are basically the same;

2) for the leakage signals that take place at different time and under different working conditions, the characteristic values collected at a same detection spot are basically the same;

3) the maximum characteristic values of normal signals are smaller than the characteristic values of leakage signals or at least most characteristic values satisfy this standard. Therefore, in characteristic extraction, the characteristic values of normal signals and of leakage signals have their respective data range and the differences in values are large and they stay stable in their own range. According to these standards and the result of data analysis, this article chooses two characteristic descriptive pressure wave signals, namely the energy characteristics and the average statistical characteristics of IMFs from 4th to 8th level:

mean variance, mean standard deviation, mean range and mean quartile range.

4 Study of Signal Recognition

Whether the selected signal characteristics can effectively represent the actual signals needs to be verified by field measurement. This article uses SVM to select characteristics and to carry out leakage identification forecast for the leakage signals after EMD and de-noising.

4.1 SIGNAL RECOGNITION EXPERIMENT BASED ON SVM

This article uses C-SVM method (suitable for small samples) with LibSVM tool cabinet. It chooses RBF kernel functions to build models for the 2 characteristics of 40 pairs of field-measured data and optimizes with grid method. Then it applies the model to categorize the 10 pairs of test data with 100% categorization accuracy. It is shown in Table 3.

TABLE 3 Recognition Result of Different Characteristics

Selected Characteristic	c	g	Categorization Accuracy of Training Set	Categorization Accuracy of Testing Set
Energy Characteristic	0.1895	3.0314	100%	100%
Statistical Characteristic	0.1895	1.7411	100%	100%

4.2 LOCATING EXPERIMENT OF SIGNAL LEAKAGE WITH EMD AND DE-NOISING

The experimental pipeline has a length of $L(km)$. If a leakage occurs $X(km)$ from the upper stream, the negative pressure wave spreads to the upper and the lower stream with a speed of $v(m/s)$. The time difference to reach the upper and the lower stream is Δt. So the pipeline leakage locating formula is as below:

$$X = \frac{1}{2}(L - v \times \Delta t),\qquad(5)$$

where the method of maximum auto-correlation of signals in the upper and lower streams can get Δt. The pipeline is 4.2 km with dynamic pressure transmitters at both sides of

the pipeline to collect instant pressure signals. The interval of sampling signals is 20 ms with sampling frequency of 50Hz. Each sampling takes 1 minute. Total sampling points are 3000. To ensure the completeness of the signals, the experiment combines the previous one-minute data with the present one to get an observation signal. So each signal is 2 minutes long and in total there are 6000 sampling points. The pressure wave spreads in the petroleum medium at a speed of 456m/s. Then, carry out correlation analysis of both pressure signals gained from the upper and lower streams of the pipeline and the EMD decomposed and re-composed signals to gain different Δt. Then, use the Equation (5) to calculate the position of leakage. The result is shown in Table 4. The positioning by EMD and de-noising signals is more accurate.

TABLE 4 Leakage Positioning Result

Experiment No.	Actual Leaking Point (km)	Original Signals			EMD and De-noising Signals		
		Δt (ms)	Positioning (km)	Positioning Error (km)	Δt (ms)	Positioning (km)	Positioning Error (km)
1	2	10	2.0977	0.0977	482	1.9901	0.0099
2	2	98	2.0777	0.0777	474	1.9919	0.0081
3	1	108	2.0754	1.0754	4226	1.1365	0.1365
4	1	8	2.0982	1.0982	1476	1.7635	0.7635
5	1	6	2.0986	1.0986	1490	1.7603	0.7603

5 Conclusions

This article is based on research data from an oil field measured dynamic pressure signals and comes to the following conclusions:

1) EMD method is applicable to the de-noising of dynamic pressure signals in petroleum pipelines.

2) The field-measured data of a pipeline can re-compose signals from 4th to 8th level.

3) EMD and de-noising signals perform better to position the leakage than original signals.

4) The energy characteristics of IMFs from 4th to 8th level can effectively represent signals;

5) The mean statistical characteristics, including mean variance, mean standard deviation, mean range and mean quartile range can effectively represent signals.

6) C-SVM method can effectively recognize dynamic pressure signals.

Acknowledgements

References

[1] You B, Peng T, Pu S 2008 The detection methods in pipeline leakage and evaluation *Oil-Gasfield Surface Engineering* **11**(27) 75-6 *(in Chinese)*

[2] Jin S 1998 The leakage positioning technology of petroleum pipeline *Acta Petrolei Sinica*, **19**(3) 93-7 *(in Chinese)*

[3] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu, H H 1998 The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis *Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences* **454** 903-95

[4] Flandrin P, Concalves P 2004 Empirical mode decompositions as a data driven wavelet-like expansions *International Journal of Wavelets, Multiresolution and Information Processing* **2**(4) 447-96

[5] Yang C 2012 Pipeline leakage detection method based on double pressure sensor and EMD *Silicon Valley* **7** 115-117 *(in Chinese)*

[6] Zhang Y, Chen S, Li J, Zhuge J, Jin S 2009 The leakage detection and positioning system of petroleum pipeline based on dynamic pressure transmitter *Chinese Journal of Structural Chemistry* **9**(22) 1347-50 *(in Chinese)*

[7] Zhang Y, Jin S, He J, Cheng S, Li J 2010 A study of characteristics extraction of pipeline leakage based on dynamic pressure signals *Acta Petrolei Sinica* **31**(2) 338-42 *(in Chinese)*

[8] Mao H, Li W, Feng X 2010 A study of pipeline leakage positioning based on EMD and relevant analysis *Science Technology and Engineering* **10**(19) 4628-32 *(in Chinese)*

[9] Hu H, Ma X 2006 The de-nosing calculation method of signals based on local wave decomposition *Journal of agricultural machinery* **1**(37) 118-20 *(in Chinese)*

[10] Lin W, Chen P, Sun J 2008 The characteristics extraction of pipeline leakage signals with adaptivity to all work conditions *Journal of Chemical Engineering* **59**(7) 1715-20 *(in Chinese)*

[11] Chang C, Lin C 2001 LIBSVM: a library for support vector machines Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Authors

Liu Wei, born in 1976, Shaanxi, China

Current position, grades: lecturer at the Department of Computer Science and Technology, Shaanxi Xueqian Normal University.
University studies: PhD at the School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology.
Scientific interest: Oil Pipeline Leak Detection Technology

Liu Hongzhao, born in 1954, Shaanxi, China

Current position, grades: Professor and PhD Candidate Supervisor.
University studies: School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology.
Scientific interest: electromechanical systems modelling, simulation, testing and control; mechanical CAD, robotics and institutional learning; mechanical dynamics and control.

Dynamical behaviour and coupling synchronization of oscillatory activities in a cell system of neural network

Kaijun Wu[1]*, Tianqi Luo[1], Huaiwei Lu[2], Yazhou Shan[1]

[1]*School of Electronic and Information Engineering, LanZhou JiaoTong Unversity, Lanzhou 730070, China*

[2]*School of Mathematics and Physics, LanZhou JiaoTong Unversity, Lanzhou 730070, China*

Received 21 April 2014, www.cmnt.lv

Abstract

The paper mainly described a mathematical model of calcium ion oscillation of non-excitable cells. Based on the model, complex oscillations caused by variations of bifurcation parameters were analysed in detail and effects of variation of parameters to synchronization were discussed when the coupling intensity of two gaps-junction- coupled calcium ion cells was certain. Through numerical simulation, under less stiffness of coupling and among a certain scope of parameters, the phenomenon that the larger the parameters were, the easier the system occurred approximate synchronization were further illustrated.

Keywords: cluster oscillation, bifurcation, fast-slow dynamics, synchronization

1 Introduction

Calcium is the most important second messenger used to activate various cellular processes relaying information within cells to regulate their activity. The calcium signal is a transient increase of the intracellular concentration [1-3], the increase of Ca^{2+} dues to entry through the cell membrane and release from internal storage compartments especially the endoplasmic reticulum (ER) and the sarcoplasmic reticulum, which results in many cells to the formation of spatiotemporal signals in form of waves of high Ca^{2+} concentration travelling across the cell and global oscillations. The information transmitted by these signals arrives as a stimulus at the plasma membrane and is translated into intracellular Ca^{2+} oscillations.

A lot of researches show that many cell functions are closely related to the changes of intracellular calcium concentration [4-6], such as muscle excitation contraction coupling, neurotransmitter release, maintenance of cell membrane excitability, regulation membrane on various ion permeability and hormone secretion. The major intracellular calcium libraries have IP3 Ca^{2+} library, Ryanodine Ca^{2+} library, Ca^{2+} library of interactions between endoplasmic reticulum, mitochondria and so on. Many scholars [7-14] established the model of the calcium oscillation and studied the calcium oscillation and synchronous behaviour in nonlinear dynamics method, revealed the oscillation complex dynamic behaviour of the biological cell system, and found that the oscillation mode, such as cycle, quasi-periodical, chaotic and integer times, the model can well simulate the oscillations of intracellular calcium, so get the very extensive research.

Synchronization of coupled systems is a new research field of nonlinear dynamics, coupling synchronization is a basic phenomenon of nonlinear dynamics. Synchronization of coupled neurons system has been a preliminary study, for example, Zheng Y H [15] have studied synchronization in ring coupled chaotic neurons with time delay, Ge M L [16] have studied synchronization oscillatory on the electrically coupled chay neurons. Studies show that Ca^{2+} transfers from a cell to another cell, intercellular coupling make multicellular calcium oscillations perform synchronization. So the study of synchronous calcium oscillations behaviour in biological cells under the reasonable parameters in changing conditions has important significance. In the paper, dynamical behaviour and coupling synchronization of oscillatory activities in a parameter changing range is analysed and the effect of synchronization of the two gaps-junction-coupled calcium ion cells is discussed on the basis of Borghans model [4].

2 Dynamical model

To study the dynamical behaviour and coupling synchronization of oscillatory activities and the effect of synchronization of the two gaps-junction-coupled calcium ion cells, the model proposed by Borghans et al [4] was used in the paper. The model consists of three basic model compartments, which is the cytosol, the endoplasmic reticulum (ER) and the mitochondria (for details see reference [6]). Consequently, the three main variables are: free Ca^{2+} concentration in the cytosol (Ca_{cyt}), free Ca^{2+} concentration in the ER (Ca_{er}), and concentration of the inositol 1, 4, 5-trisphosphate (IP_3). The evolution of the model system is descried by the following differentials. The Ca_{cyt} changes due to the influx of extracellular Ca^{2+} (J_{in}), the passive efflux of Ca^{2+} changes from the cytosol

* *Corresponding author* e-mail: wkj@mail.lzjtu.cn

into the extracellular medium (J_{out}), and from the intracellular Ca^2 store into the cytosol (J_{er}). Moreover, Ca^2 is pumped into (J_{pump}) and released from (J_{leak}) the intracellular store. The evolution of the model system is descried by the following differential equations:

$$\frac{dCa_{cyt}}{dt} = J_{in} + J_{leak} - J_{pump} + J_{er} - J_{out} , \tag{1}$$

$$\frac{dIP_3}{dt} = J_A - J_D - J_C , \tag{2}$$

$$\frac{dCa_{er}}{dt} = J_{pump} - J_{leak} - J_{er} , \tag{3}$$

where:

$$J_{in} = k_{in1}r + k_{in2} , \tag{4}$$

$$J_{leak} = rk_{leak} \frac{IP_3{}^4}{IP_3{}^4 + k_a{}^4} \frac{Ca_{er}{}^2}{Ca_{er}{}^2 + k_y{}^2} \frac{Ca_{cyt}{}^4}{Ca_{cyt}{}^4 + k_z{}^4} , \tag{5}$$

$$J_{pump} = k_{pump} \frac{Ca_{cyt}{}^2}{Ca_{cyt}{}^2 + k_2{}^2} , \tag{6}$$

$$J_{out} = k_{out}Ca_{cyt} , \tag{7}$$

$$J_{er} = k_f Ca_{er} , \tag{8}$$

$$J_A = rk_p , \tag{9}$$

$$J_C = \varepsilon IP_3 , \tag{10}$$

$$J_D = k_d \frac{IP_3{}^2}{IP_3{}^2 + K_p{}^2} \frac{Ca_{cyt}{}^4}{Ca_{cyt}{}^4 + K_d{}^4} . \tag{11}$$

In the paper, we take k_{pump} as parameter to consider the influence of the oscillation to the model, other parameters are listed in the Table 1, among of them, the unit of k_{leak}, k_{in1}, k_{in2}, k_p and k_d is μMmin^{-1} and the unit of k_f and k_{out} is min^{-1}, the unit of k_2, k_a, k_y, k_z, K_d and K_p is μM.

TABLE 1 Model parameters for all calculated results unless stated

parameter	k_{leak}	k_{in1}	k_{in2}	k_p	k_d	k_f	k_{out}	k_2
value	19.5	1.0	2.0	2.5	80	1.0	10	0.1
parameter	k_a	k_y	k_z	K_d	K_p	r	ε	
value	0.4	0.2	0.3	0.4	1.0	1.12	0.1	

3 Analysis of corresponding dynamical behaviour

3.1 BIFURCATION ANALYSIS OF FAST AND SLOW DYNAMICS

Regarding of the system, we use method of fast and slow dynamics. Take the third equation as tardyon system and other equations as tachyon systems, which mean that Ca_{er} works as slow parameter and studies the model through changing the value of k_{pump}. When $k_{pump} = 6.6$ the system

appears the cluster oscillation pattern shown in the Figure 1a. We use the fast and slow dynamics fork graph in Figure 1b to analyze its dynamics characteristics and type of the cluster oscillation.
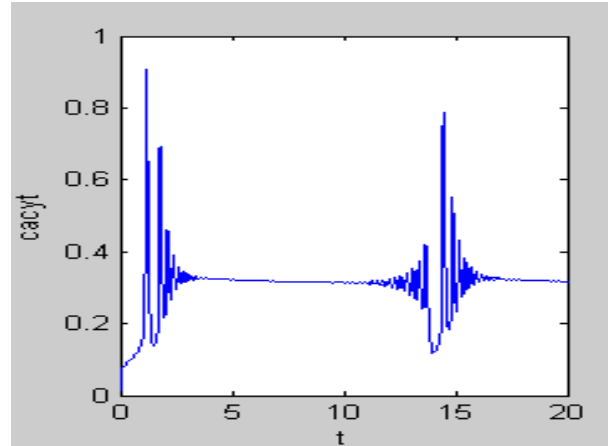
FIGURE 1a the time series of busting oscillations of the whole system
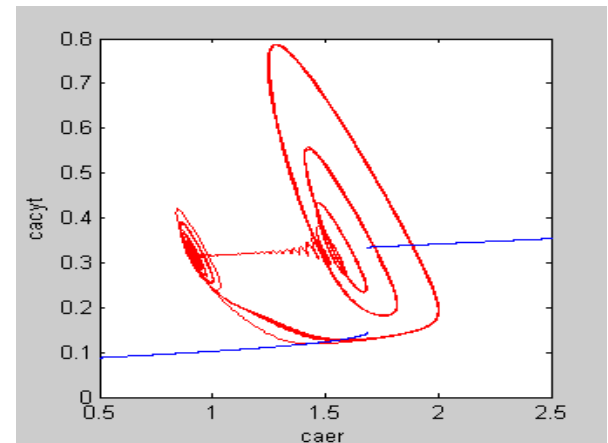
FIGURE 1b the bifurcation diagram of fast subsystem with Ca_{er} as the bifurcation parameter, the red closed trajectories represents for the phase diagram of (Ca_{er}, Ca_{cyt})

The balanced point of fast parameter system along with the change of its slow parameter Ca_{er} forms a resembling S-shape bifurcation line (the part between two full lines is the unstable part) on (Ca_{er}, Ca_{cyt}) plane in Figure 1b, and Figure 1a adds the corresponding Ca_{cyt} time course diagram. S-shape bifurcation line formed in fast parameter system can divide into upper, middle and below parts. In Figure 1a there exists a point-point bursting, whose main characteristic is that whatever bursting is active or resting, it all rely on the stabilize state of fast parameter system. Oscillation of point-point bursting is very commonly seen in the cytoplasm calcium oscillation model and can be well described by fast and slow dynamics.

The upper and below parts of bifurcation line of fast parameter system are made up of stable points, while the middle part consists unstable points. Observed Figure 1b in anticlockwise direction, the beginning and the ending of oscillation accompany with a folding bifurcation, and with the increasing of the time, the below part loses the stable state step by step through a folding bifurcation until reaches the stable upper part where Ca_{cyt} achieves its first

463

maximum value as spiking. Therefore, when $k_{pump} = 6.6$, cluster oscillation pattern of the model presents "fold/fold" point-point cluster oscillation. Point-ring calcium oscillation might occur in the model with the change of k_{pump} when takes Ca_{er} as the slow parameter.

3.2 OVERALL BIFURCATION ANALYSIS ON k_{pump}

When parameter k_{pump} changes in [5, 16] of system (1), we use C language and four order Runge-Kutta method with variable step size to make numerical simulation. The overall bifurcation diagram of the entire system is shown in Figure 2a. Concluded from the figure, calcium oscillation only occurs in a limited range: first it starts from a stable state to an unstable state through a hopf bifurcation and then reaches the final stable state through a hopf bifurcation. When $k_{pump} = 5.862$, system occurs hopf bifurcation to generate limit cycle and enters in the oscillation state. When $k_{pump} = 15.31$, again system enters in the stable state through a hopf bifurcation. We testify the result by observing Figure 2b. To learn the dynamic characteristic of the model perfectly, we analyse phase diagrams by setting the different parameters of k_{pump}.
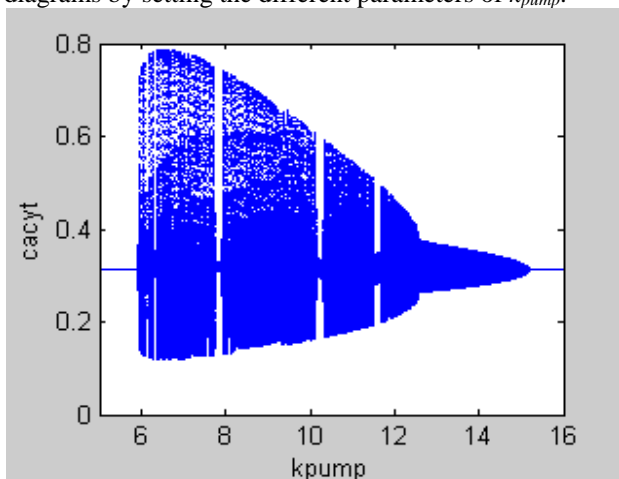
FIGURE 2a the bifurcation diagram of Ca_{cyt} versus the bifurcation parameter k_{pump}
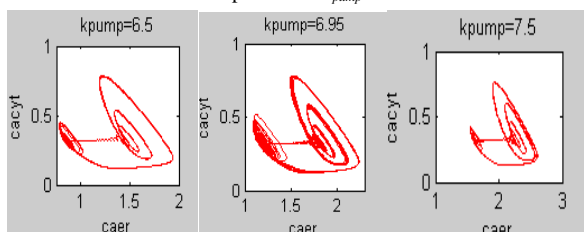
FIGURE 2b the small periodic window
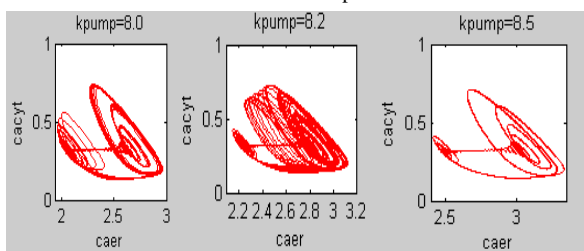
FIGURE 2c the small periodic window
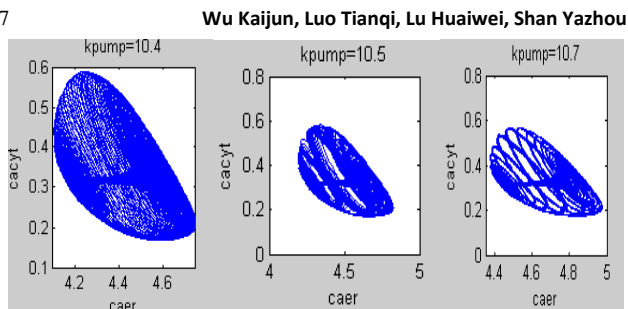
FIGURE 2d the small periodic window

With the increasing of parameter k_{pump}, oscillation system gets into the bursting state that we can observe the periodic bursting, quasi-periodic bursting and chaotic bursting. It is rarely seen in the calcium oscillation mathematics model about quasi-periodic phenomenon [8], which is able to well observe in this model with the changing of parameter k_{pump}, moreover some relatively narrow quasi-periodic windows such as Figure 2b, Figure 2c and Figure 2d can also be found. Quasi-periodic phenomenon is normally judged by calculating its corresponding Lyapunov exponents value, because when the exponent value is larger than 0, the corresponding oscillation is in chaotic state and when the value is smaller than 0, the oscillation is in stable state, however, when Lyapunov exponent equals with 0, the oscillation is in periodic state. Quasi-periodic phenomenon can also be differentiated from phase diagrams, normally a circular ring can be observed in corresponding phase diagrams. In Figure 2b system is in periodic movement when $k_{pump} = 6.5$; in quasi-periodic movement when $k_{pump} = 10.5$ for we can see a circular ring clearly in the second phase diagram of Figure 2d. When $k_{pump} = 10.4$, however, system lies in chaotic movement state. If k_{pump} continues to increase, system repeats periodic-quasi-periodic-chaotic movements.

4 Analysis of gap-junction-coupled synchronization of two calcium oscillation cells

Study of synchronization started from 1673, till now many scholars from variety fields have obtained some significant results [15-21]. Synchronization phenomenon broadly exists in biological cell system and plays important role in some basic activities of organism. Synchronization only occurs under smaller coupling intensity for chaotic system [19]. This paper mainly studies two Borghans models through junction coupling and the effect of synchronization with the changing of bifurcation under small coupling intensity. The coupling system consists with the following equations, among of them, $D(ca_{cyt2} - ca_{cyt1})$ and $D(ca_{cyt1} - ca_{cyt2})$ are coupling terms, D is coupling intensity and its value is in (0,0.1). Values of other parameters are same with the system listed in Table 1.

464

Wu Kaijun, Luo Tianqi, Lu Huaiwei, Shan Yazhou

$$\frac{dca_{cyt1}}{dt} = J_{in} + J_{leak} - J_{pump} + \tag{12}$$
$$J_{er} - J_{out} + D(ca_{cyt2} - ca_{cyt1}),$$

$$\frac{dIP_{31}}{dt} = J_A - J_D - J_C, \tag{13}$$

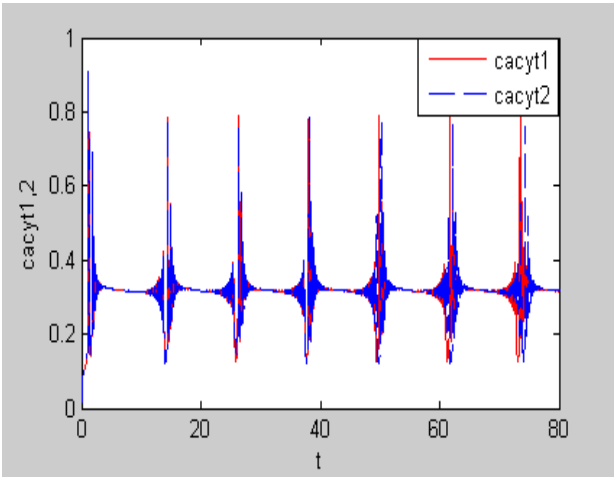$$\frac{dca_{er1}}{dt} = J_{pump} - J_{leak} - J_{er}, \tag{14}$$

$$\frac{dca_{cyt2}}{dt} = J_{in} + J_{leak} - J_{pump} + , \tag{15}$$
$$J_{er} - J_{out} + D(ca_{cyt1} - ca_{cyt2}),$$

$$\frac{dIP_{32}}{dt} = J_A - J_D - J_C, \tag{16}$$

$$\frac{dca_{er2}}{dt} = J_{pump} - J_{leak} - J_{er}. \tag{17}$$

FIGURE 3c the time series of bursting oscillations when $k_{pump} = 8$
FIGURE 3 the time series of Cacyt in the first (red solid line) and in the second (blue dotted line) coupled cell when D=0.001
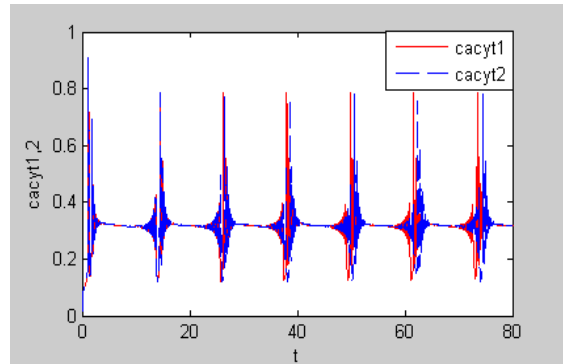
FIGURE 4a the time series of bursting oscillations when $k_{pump} = 6.6$

FIGURE 3a the time series of bursting oscillations when $k_{pump} = 6.6$
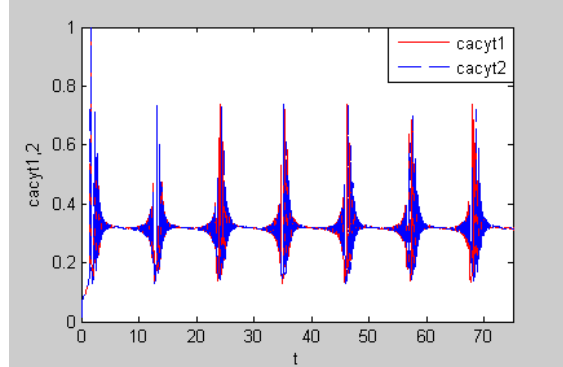
FIGURE 4b the time series of bursting oscillations when $k_{pump} = 8$

FIGURE 3b the time series of bursting oscillations when $k_{pump} = 8$
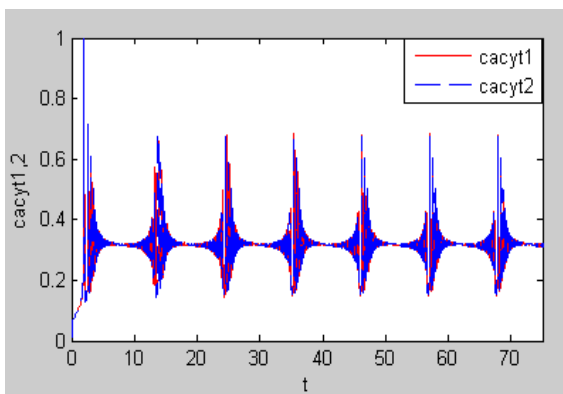
FIGURE 4c the time series of busting oscillations when $k_{pump} = 9$
FIGURE 4 the time series of Ca_{cyt} in the first (red solid line) and in the second (blue dotted line) coupled cell when D=0.003

Wu Kaijun, Luo Tianqi, Lu Huaiwei, Shan Yazhou

Seen from Figure 3, when coupling intensity is fixed ($D = 0.001$) and other parameters remains unchanged and parameter k_{pump} changes in a certain range, coupling system achieves approximate synchronization but not completely synchronization with the increasing of k_{pump}. Similarly, when $D = 0.003$, observed from Figure 4, we can conclude that, when $k_{pump} = 6.6$, two coupling system don't synchronize; when $k_{pump} = 8$, two coupling system gradually synchronize approximately; when $k_{pump} = 9$, system reaches approximate synchronization. Thus, we can conclude: when coupling intensity is confirmed, the occurrence of approximate synchronization of two gap-junction-coupled biological cells is related to the values of the parameters. In this system, in a certain range, the larger the parameters are, the easier the system is likely to occur approximate synchronization.

5 Conclusions

For Borghans model of non-excitable calcium ion oscillation, we firstly took the calcium ion concentration in endoplasmic reticulum as the slow-change parameter and mainly researched the dynamical mechanism of the "fold/fold" point-point cluster oscillation generated in the model when $k_{pump} = 6.6$ using fast and slow dynamic bifurcation method. Of course, if k_{pump} were given other values, the model might have other complex oscillation patterns. Secondly, we analysed the overall bifurcation diagram taking k_{pump} as the bifurcation parameter and got complex calcium oscillation patterns through the diagram and the corresponding phase diagrams. In some small intervals of k_{pump}, quasi-periodic movements can be

observed. For instance, the system went through the periodic, quasi-periodic and chaotic oscillation patterns orderly when the value of k_{pump} was 6.5, 8, 8.2. When k_{pump} equalled to 8.4 to 10.0, the above process repeated. Finally, the paper mainly discussed the synchronization problem of two gap-junction-coupled calcium ions. Through the analysis and numerical simulation, we found that complete synchronization could not happen. The value of k_{pump} can affect the synchronization of the coupling model when the coupling intensity is confirmed. To be specific, in the changing range of k_{pump}, the larger the value is, the closer the coupling system approach to the approximate synchronization state. These results are helpful to the understanding of the effect of k_{pump} in complex calcium oscillations.

Cluster oscillation pattern has attracted more and more concentrations of scholars for its varied cluster types, and synchronization of coupling biological cells plays a significant role in biology information transmission and coding, therefore, studying the synchronization problem of coupling cluster oscillation has a positive guiding meaning to solve some practical issues.

Acknowledgments

References

[1] Putney J, Bird G 1993 The inositol phosphate-calcium signalling system in non-excitable cells *Endocrine Reviews* **14** 610-31

[2] Berridge M, Lipp P, Bootman M 2000 The versatility and universality of calcium signalling *Nature Rev Mol Cell Biol* **1** 11-22

[3] Falcke M 2004 Reading the patterns in living cells – the physics of Ca^{2+} signalling *Advances in Physics* **53** 255-440

[4] Borghans A M, Dupont G, Goldbeter A 1997 Complex intracellular calcium oscillations *Biophys Chem* **66** 25-41

[5] Shen P, Larter R 1995 Chaos in intracellular calcium oscillations in a new model for non-excitable cells Cell Calcium **17** 225-32

[6] Marhl M, Haberichter T, Brumen M 2000 Complex calcium oscillations and the role of mitochondria and cytosolic proteins *Biosystems* **57** 75-86

[7] Grubelnik V, Larsen A Z, Kummer U, Olsen L F, Marhl M 2001 Mitochondria regulates the amplitude of simple and complex calcium oscillations *Biophysical Chemostry* **94** 59-74

[8] Perc M, Marhl M 2003 Different types of bursting calcium oscillations in non-excitable cells *Chaos, Solitons and Fractals* **18** 759-73

[9] Perc M, Marhl M 2004 Synchronization of regular and chaotic oscillations: The role of local divergence and the slow passage effect – A case study on calcium oscillations *International journal of bifurcation and chaos in applied sciences and engineering* **14**(8) 2735-51

[10] Izhikevich E M 2006 Dynamical systems in neuron science: the geometry of excitability and bursting *The MIT Press* USA

[11] Ji Q B, Lu Q S, Yang Z Q, Duan L X 2008 Bursting Ca2+oscillations and synchronization in coupled cells *Chinese Physics Letters* **25** 3879-83

[12] Nchange A K, Kepseu W D, Woafo P 2008 Noise induced intercellular propagation of calcium waves *Physica A* **387** 2519-25

[13] Zhang F, Lu Q S, Su J Z 2009 Transition in complex calcium bursting induced by IP3 degradation *Chaos, Solitons & Fractals* **5** 2285-90

[14] Aguilar-López R, Martínez-Guerra R, Puebla H, Hernández-Suárez R 2010 High order sliding-mode dynamic control for chaotic intracellular calcium oscillations *Nonlinear Analysis: Real World Applications* **11** 217-31

[15] Zheng Y H, Lu Q S 2008 Synchronization in ring coupled chaotic neurons with time delay *Journal of dynamics and control* **6**(3) 208-12

[16] Ge M L, Guo H Y, Wang G J, Yan W L 2003 Research for synchronization oscillatory on the electrically coupled chay neurons *Acta biophysica sinica* **19**(2) 135-40

[17] Chandrasekaran L, Achuthan S, Canavier C 2011 Stability of two cluster solutions in pulse coupled networks of neural oscillators *Journal of Computational Neuroscience* **30**(2) 427-45

[18] Schöll E, Selivanov A, Lehnert J, Dahms T, Hövel P, Fradkov A 2012 Control of synchronization in delay-coupled networks *International Journal of Modern Physics B* **26**(25) 1246-57

[19] Wang Q Y, Duan Z S, Feng Z S, Chen G, Lu Q 2008 Synchronization transition in gap-junction-coupled leech neurons Physica A, **387**(16) 4404-10

[20] Shawn D, Bard G, Nathaniel N 2008 Intrinsic heterogeneity in oscillatory dynamics limits correlation-induced neural synchronization *J Neurophysiol* **108** 2115–33

[21] Witte H, Putsche P, Eiselt M, Schwab K, Wacker M, Leistritz L 2011 Time-variant analysis of phase couplings and amplitude–frequency dependencies of and between frequency components of EEG burst patterns in full-term newborns *Clinical Neurophysiology* **122** 253-66

Authors

Kaijun Wu, born in March, 1978, Shandong, China

Current position, grades: associate professor and master tutor in Lanzhou Jiaotong University.
University studies: PhD student of automation and electrical engineering, Lanzhou Jiaotong University, P.R. China.
Scientific interests: nonlinear dynamics, neural networks.
Publications: 35 papers, 2 textbooks.

Tianqi Luo, born in February, 1993, Gansu, China

Current position, grades: undergraduate student in Lanzhou Jiaotong University.
University studies: Bachelor degree student in the school of Electronic and Information Engineering, Lanzhou Jiaotong University.
Scientific interests: nonlinear dynamics, neural networks.

Huaiwei Lu, born in March, 1959, Gansu, China

Current position, grades: professor and doctoral tutor in Lanzhou Jiaotong University.
University studies: master degree in the University of electronic science and technology of China in 1989.
Scientific interests: optical fiber communications, optical fiber passive components.
Publications: 48 papers.

Yazhou Shan, born in June, 1990, Henan, China

Current position, grades: master student in Lanzhou Jiaotong University.
University studies: master degree student in the school of Electronic and Information Engineering, Lanzhou Jiaotong University.
Scientific interests: nonlinear dynamics, neural network.

Finite time control for probe descent near small bodies

Chunhun Liang[1, 2], Yuanchun Li[1, 3*]

[1]*College of Communication Engineering, Jilin University, 130012, Changchun, China*

[2]*College of Electrical and Information, Changchun Institute of Technology, 130012, Changchun, China*

[3]*Department of Control Engineering, Changchun University of Technology, 130012, Changchun, China*

Abstract

For the purpose of probe soft landing on small bodies safely, this paper focuses on the orbital dynamics and describes a new control algorithm for the probe descent near a rotating small body. The general formulation of the probe equations of motion in the Body-Fixed Coordinate system is obtained through Newton's second law firstly. Then the nominal polynomial trajectory of vertical direction is planned in the condition of fuel consumption suboptimum. Considering uncertainties and perturbations, the control laws based on Adaptive Terminal Sliding Mode with compensation term are developed to track the desired trajectory finally. Suppose the initial conditions presented in this paper, position and velocity errors tend to zero in the finite time in the phase of sliding mode motion so it will make the task succeed during descent phase near small bodies. Finally, the effectiveness of the guidance and control algorithm is verified by MATLAB simulations. The proposed algorithm can fast and accurately track the planned trajectory in the finite time and is robust to parameter uncertainty, feedback state error and external disturbances. The validity is conformed by computer simulation.

Keywords: adaptive terminal sliding mode control, soft landing, descent near small body, guidance and control

1 Introduction

In recent years our solar system is small bodies (e.g., mainly asteroids and comets) have received increased attention because of the insight they can give about the history of the solar system. In this respect, several studies have been conducted and some missions have been flown and planned [1]. The missions flown include: On 12 February 2001, Near Earth Asteroid Rendezvous (NEAR) Shoemaker became the first spacecraft landing on a small body 433 Eros after a busy 4-year cruise [2-4]. ISAS had planed to launch a sample and return spacecraft MUSES-C in 2002 and the spacecraft arrived at near earth asteroid 1989ML in 2003 [5, 6]. And Rosetta - ESA Comet Mission, will fly by asteroids Steins and Lutetia in 2004.

The small size, irregular shape, mass distribution as well as the state of rotation of an asteroid, have significant effects on the evolution of the orbit of the spacecraft. The high precise guidance, navigation, and control (GNC) system is needed for safe and precise asteroid landing [7]. Several papers have addressed the problems of orbital dynamics, autonomous guidance and control for landing on small body about asteroids. PD and Sliding Mode Controller are designed to tracking the fuel consumption suboptimal nominal trajectory [8]. The authors provide scope for precision landing of payloads on minor solar system bodies using a probe as a reference to generate a line of sight to the landing location [9, 10]. In these methods the probe commanded acceleration is collinear with the vehicle velocity, and the spacecraft trajectories are restricted to a plane. With gravitational effects taken into account, a new solution to the fuel optimal vertical landing on an asteroid was obtained. The open-loop guidance based on a pseudo way-point generation algorithm is provided that uses a discrete linear time-varying model of the dynamics that incorporates required thruster silent times. Then it is transferred to second-order cone programming problem. Feedback control is implemented to track the pseudo way-point trajectory [11]. Since the gravitation on an asteroid is low, a wheeled vehicle would likely bounce back from hitting the surface, and be difficult to control. A first order model of the dynamics of hopping robots is developed to estimate the total time and distance covered from an initial bounce to a stop due to friction and restitution coefficients. Then sliding-mode control is applied to discrete formation control [12].

The rapidity and robustness of tracking the desired trajectory are not ideal in condition of uncertainties and perturbations for normal Sliding Mode Control. The Terminal Sliding Mode Control with compensation term is proposed to deal with MIMO system [13, 14]. The nonlinear term is introduced into the design of the sliding mode surface such that the state variable of the sliding mode tends to zero in the finite time in the phase of sliding mode motion.

The body of this paper consists of four main sections followed by a conclusion. Section 2 describes the probe descent orbital dynamic model through Newton's second law referring to lunar soft landing as well as the irregular gravitation potential model of small body through harmonic series expansion method. The nominal trajectories considering fuel suboptimal are designed in

---

* *Corresponding author* e-mail: liyc@mail.ccut.edu.cn

section 3. In section 4, an Adaptive Terminal Sliding Mode Controller with compensation term is designed to track the desired trajectory so the probe can land on the surface of small body quickly and safely while the boundary of uncertainties and perturbations are unknown. Computer simulations are carried out to illustrate the effectiveness and robustness of the control laws.

2 Small body and probe models

2.1 PROBE ORBITAL DYNAMICS MODEL

The probe orbital dynamic model is built in the three dimensional coordinates referring to model of soft landing on lunar [15], as shown in Figure 1. The small body has the instantaneous rotation angular velocity vector of ω. Affected by the irregular and weak gravitational potential, the perturbation force accelerations such as solar radiation pressure and solar gravitational, the probe cannot always retain the same plane during landing.
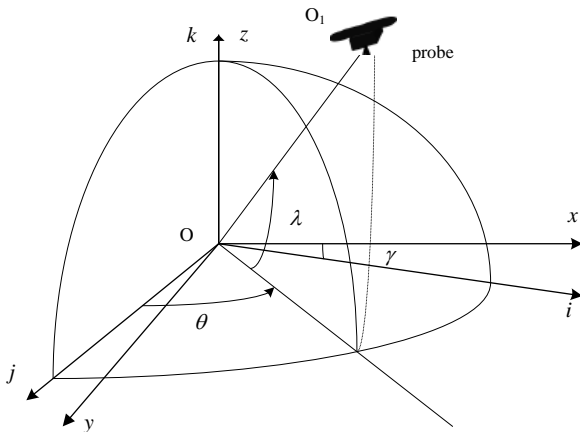
FIGURE 1 Viewing geometry of coordinate systems

For further understanding of equations of motion, it is convenient to define the following coordinate frames. The small body Central Inertial Coordinate system $oijk$ with the body equator as the reference plane, coordinate system origin o is the centre of mass of the small body. The asteroid Body-Fixed Coordinate system $oxyz$ is defined. Let the coordinate system origin o is fixed with the asteroid mass centre, and the z-axis coinciding with the asteroid's maximum moment of inertia, the x-axis and the y-axis coinciding with the minimum and intermediate moments of inertia, respectively. For simplicity, we assume that the spin axis of the target asteroid is coincided with the asteroid's maximum moment of inertia. All the coordinates are satisfied with right handed systems. Without lose of generality, we assume that $oijk$ and $oxyz$ coincide with each other at the beginning of descending and soft landing. And o_1 is the centre of mass of the probe. Supposed the small body is a unit mass and the fuel consumption is omitted, system orbital dynamic equations in the Body Central Inertial Coordinate can be derived based on Newton's second law. The detail process is presented as below:

$$\left[\frac{d\vec{V}}{dt}\right]_{ijk} = P_1 \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}_{ijk} + \vec{G}_{ijk} + \vec{\Delta}_{ijk} , \tag{1}$$

$$P_1 = \begin{bmatrix} \cos\lambda\cos\theta & \sin\theta & -\sin\lambda\cos\theta \\ -\cos\lambda\cos\theta & \cos\theta & \sin\lambda\sin\theta \\ \sin\lambda & 0 & \cos\lambda \end{bmatrix},$$

where $\vec{V}$ and U are velocity vector and gravity function of probe in $oijk$ coordinate, and Δ is perturbation force acceleration which includes solar radiation pressure perturbation and solar gravitational perturbation. u_x, u_y, u_z are components of control acceleration, and u_z coinciding with the vector from the small body mass centre to the probe, u_x perpendicular with u_z and pointing to the motion of probe, u_x, u_y and u_z complied with the right handed system. We define position and velocity vector of probe $\vec{R}$ and $\vec{V}_L$ in the Body-Fixed Coordinate, and $\vec{\omega}$ is the Body-Fixed Coordinate system $oxyz$ rotating velocity relative to Central Inertial Coordinate $oijk$.

The following equation is derived based on Koch law:

$$\vec{V} = \vec{V}_L + \vec{\omega} \times \vec{R} . \tag{2}$$

Taking the derivative of the Equations (2), the acceleration of probe in the Body Central Inertial Coordinate is deduced as Equation (3):

$$\left[\frac{d\vec{V}}{dt}\right]_{ijk} = \left[\frac{d\vec{V}_L}{dt}\right]_{ijk} + \vec{\omega}\times\left[\frac{d\vec{R}}{dt}\right]_{ijk} + \left[\frac{d\vec{\omega}}{dt}\right]_{ijk}\times\vec{R} . \tag{3}$$

Similarly, Equation (4) is derived based on Koch law

$$\left[\frac{d\vec{V}_L}{dt}\right]_{ijk} = \left[\frac{d\vec{V}_L}{dt}\right]_{xyz} + \vec{\omega}\times\vec{V}_L$$
$$\left[\frac{d\vec{R}}{dt}\right]_{ijk} = \vec{V}_L + \vec{\omega}\times R \tag{4}$$

Suppose the small body rotation speed is uniform with the constant rate of revolution ω:

$$\left[\frac{d\vec{\omega}}{dt}\right]_{xyz} = 0 . \tag{5}$$

Substituting Equation (4) and Equation (5) into Equation (3):

$$\left[\frac{d\vec{V}_L}{dt}\right]_{xyz} = \left[\frac{d\vec{V}}{dt}\right]_{ijk} - 2\vec{\omega}\times\vec{V}_L - \vec{\omega}\times\left(\vec{\omega}\times\vec{R}\right) . \tag{6}$$

Therefore, the acceleration of probe is in the Body-Fixed Coordinate:

$$\left[\frac{d\vec{V}}{dt}\right]_{xyz} = P_1 P_2 \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} + \vec{G}_L - 2\vec{\omega}\times\vec{V}_L - \vec{\omega}\times\left(\vec{\omega}\times\vec{R}\right), \qquad (7)$$

$$P_2 = \begin{bmatrix} \cos\gamma & 0 & -\sin\gamma \\ 0 & 1 & 0 \\ \sin\gamma & 0 & \cos\gamma \end{bmatrix}.$$

Small body gravitation acceleration projection formation in the Body-Fixed Coordinate is indicated as:

$$\vec{U}_L = \vec{G}_L - \vec{\omega}\times(\vec{\omega}\times\vec{R}).$$

Substituting above equation into Equation (3),

$$\begin{bmatrix} \dot{V}_{xL} \\ \dot{V}_{yL} \\ \dot{V}_{zL} \end{bmatrix} = P_1 P_2 \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} - \begin{bmatrix} U_{xL} \\ U_{yL} \\ U_{zL} \end{bmatrix} - \begin{bmatrix} 2\omega_L V_{zL} \\ 0 \\ -2\omega_L V_{xL} \end{bmatrix}.$$

V_{xL}, V_{yL} and V_{zL} are probe velocity projection in the Body-Fixed Coordinate, ω_L is the rotation rate of small body, U_{xL}, U_{yL}, U_{zL} are gravity acceleration projection in the Body-Fixed Coordinate. Quadratic component is omit due to ω_L is very small, then the orbital dynamics of the probe in the Body-Fixed Coordinate is:

$$\begin{aligned} &\dot{x}_L = V_{xL}, \dot{y}_L = V_{yL}, \dot{z}_L = V_{zL}, \\ &\dot{V}_{xL} = u_{xL} - U_{xL} + 2\omega V_{yL} + \Delta_{xL}, \\ &\dot{V}_{yL} = u_{yL} - U_{yL} - 2\omega V_{xL} + \Delta_{yL}, \\ &\dot{V}_{zL} = u_{zL} - U_{zL} + \Delta_{zL} \end{aligned} \qquad (8)$$

where x_L, y_L, z_L are probe velocity projections in the Body-Fixed Coordinate.

$$U_{xL} = \frac{\partial V(R)}{\partial x}, \ U_{yL} = \frac{\partial V(R)}{\partial y}, \ U_{zL} = \frac{\partial V(R)}{\partial z}.$$

$V(R)$ indicates gravitational potential function, which is expressed using spherical harmonics. And:

$$\begin{bmatrix} u_{xL} & u_{yL} & u_{zL} \end{bmatrix}^T = P_1 P_2 \begin{bmatrix} u_x & u_y & u_z \end{bmatrix}^T,$$

is the component of acceleration in the Body-Fixed Coordinate. Δ_{xL}, Δ_{yL} and Δ_{zL} are perturbation force accelerations such as solar radiation pressure and solar gravitational.

2.2 GRAVITY FIELD MODEL OF SMALL BODY

The two main methods calculating the gravitational potential include harmonic series expansion method and polyhedral method. To simplify the calculation, the second method which has the analytical form is adopted in this paper. Considering the special performance during probe

descent and landing, we assume the small body is a homogeneous tri-axial ellipsoid with axes a, b, c. The gravitational potential can be expanded into a series of spherical harmonics [16], that is:

$$V(R) = \frac{GM}{R}\left\{ 1 + \sum_{n=1}^{\infty}\sum_{m=0}^{n}\left(\frac{R_0}{R}\right)^m P_{nm}(\sin\theta) \atop \times[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda] \right\}, \quad (9)$$

where GM equals to the product of the gravitational constant and the mass of small body, n is the degree and m is the order. R_0 is the reference radius(generally the largest equatorial radius), R, θ and λ are the distance from the mass centre of small body to the spacecraft, the latitude and longitude respectively in a Body-Fixed Coordinate system whose origin is at the centre of mass of the body. P_{nm} is the associated Legendre polynomials, C_{nm} and S_{nm} are coefficients of the potential determined by the mass distribution with in the small body. We can calculate the coefficients taking into account the above suppose, for such a symmetric body. $S_{nm} = 0$ for all n or m, $C_{nm} = 0$ for n or m odd and while other conditions:

$$C_{nm} = \frac{3}{R_0^{\ n}}\frac{(n/2)!(n-m)!}{2^m(n+3)(n+1)!}(2-\delta_{0m})\times$$

$$\sum_{i=0}^{int(\frac{n-m}{4})} \frac{(a^2-b^2)^{\frac{m+4i}{2}}\left[c^2-\frac{1}{2}(a^2+b^2)\right]^{\frac{n-m-4i}{2}}}{16^i(\frac{n-m-4i}{2})!(\frac{m+2i}{2})!i!},$$

where δ_{0m} is Kronecker symbol, and the value of δ_{0m} is

$$\delta_{0m} = \begin{cases} 0, m = 0 \\ 1, m = 1 \end{cases}.$$

For our purposes we have stopped the expansion of Equation (9) to the second order, so we get the following coefficient:

$$C_{20} = \frac{2c^2-(a^2+b^2)}{10R_0^{\ 2}}, C_{22} = \frac{a^2-b^2}{20R_0^{\ 2}}.$$

So, components of small body gravity are obtained as follow:

$$U_{x_L} = \frac{\partial V(R)}{\partial x_L} =$$

$$-\frac{GMx_L}{R^3}\begin{bmatrix} 1 + \frac{3}{2}C_{20}\left(\frac{R_0}{R}\right)^2\left(5\frac{z_L^{\ 2}}{R^2}-1\right) \\ +3C_{22}\left(\frac{R_0}{R}\right)^2\left(5\frac{x_L^{\ 2}-y_L^{\ 2}}{R^2}-2\right) \end{bmatrix}, \quad (10)$$

$$U_{y_L} = \frac{\partial V(R)}{\partial y_L} =$$

$$-\frac{GMy_L}{R^3}\left[\begin{array}{l} 1 + \frac{3}{2}C_{20}\left(\frac{R_0}{R}\right)^2\left(5\frac{z_L^2}{R^2}-1\right) \\ +3C_{22}\left(\frac{R_0}{R}\right)^2\left(5\frac{x_L^2-y_L^2}{R^2}+2\right) \end{array} \right], \quad (11)$$

$$U_{z_L} = \frac{\partial V(R)}{\partial z_L} =$$

$$-\frac{GMz_L}{R^3}\left[\begin{array}{l} 1 + \frac{3}{2}C_{20}\left(\frac{R_0}{R}\right)^2\left(5\frac{z_L^2}{R^2}-3\right) \\ +15C_{22}\left(\frac{R_0}{R}\right)^2\left(\frac{x_L^2-y_L^2}{R^2}\right) \end{array} \right], \quad (12)$$

$$R = \sqrt{x_L^2 + y_L^2 + z_L^2}$$

3 Guidance and control

3.1 NOMINAL TRAJECTORY GUIDANCE

The desired descent altitude and velocity is planned in order to satisfy the requirements of soft landing on the surface of small bodies**.** We suppose the nominal trajectory is fuel suboptimal polynomial trajectory employed by Apollo landing lunar. The acceleration of one axis is planned as $a = c_0 + c_1 t$, and then desired descent altitude and velocity are shown after the integral.

$$v = c_0\left(t-t_0\right) + \frac{1}{2}c_1\left(t-t_0\right)^2 + v_0, \quad (13)$$

$$r = \frac{1}{2}c_0\left(t-t_0\right)^2 + \frac{1}{6}c_1\left(t-t_0\right)^3 + v_0(t-t_0) + r_0. \quad (14)$$

The undetermined parameters c_0 and c_1 are obtained if the boundary condition is given by on the condition of the limited descend time.

$$c_0 = -\frac{2v_t + 4v_0}{t_{go}} + 6\frac{r_t - r_0}{t_{go}^2}, (15)$$

$$c_1 = \frac{6v_t + 6v_0}{t_{go}^2} - 12\frac{r_t - r_0}{t_{go}^3}, \quad (16)$$

The parameters with subscript t stand for final states of the probe, and the parameters with subscript 0 stand for initial states of the probe, t_{go} is the descent time.

3.2 TERMINAL SLIDING MODE CONTROL WITH COMPENSATION TERM

Substituting $x_{11} = x_L$, $x_{12} = V_{xL}$, $x_{21} = y_L$, $x_{22} = V_{yL}$, $x_{31} = z_L$ and $x_{32} = V_{zL}$ into Equation (8):

$$\begin{cases} \dot{x}_{11} = x_{12} \\ \dot{x}_{12} = u_{xL} - U_{xL} + 2\omega_L x_{22} + \Delta_{xL} \\ \dot{x}_{21} = x_{22} \\ \dot{x}_{22} = u_{yL} - U_{yL} - 2\omega_L x_{12} + \Delta_{yL} \\ \dot{x}_{31} = x_{32} \\ \dot{x}_{32} = u_{zL} - U_{zL} + \Delta_{zL} \end{cases}, \quad (17)$$

$$f(X,t) = \begin{bmatrix} -U_{xL} + 2\omega_L x_{22} \\ -U_{yL} - 2\omega_L x_{12} \\ -U_{zL} \end{bmatrix}, \ b(X,t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Modelling uncertainties related to irregular gravitational potential and perturbations as below:

$$\Delta f(x) = \begin{bmatrix} 0.2\sin(\omega_L t)\cdot U_{xL} \\ 0.2\sin(\omega_L t)\cdot U_{yL} \\ 0.2\sin(\omega_L t)\cdot U_{zL} \end{bmatrix}, \ d(t) = \begin{bmatrix} 0.02\sin(2\pi t) \\ 0.02\sin(2\pi t) \\ 0.02\sin(2\pi t) \end{bmatrix}.$$

So the model is similar with the following nonlinear system:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(X,t) + \Delta f(X,t) + b(X,t)u + d(t). \end{cases} \quad (18)$$
$$X = [x_1^T \ x_2^T]^T = [x_1^T \ \dot{x}_1^T]^T$$

Suppose: Uncertainty $\Delta f(X,t)$ and external disturbance $d(t)$ are bounded and satisfy $|\Delta f(X,t)| \le F(X,t)$, $|d(t)| \le D(t)$, $F(X,t)$ and $D(t)$ are nonnegative functions. Definition error vector:

$$E = X - X_d = [e^T \ \dot{e}^T]^T, \ e = x_i - x_{id} = [e_1 \ e_2 \ e_3]^T.$$

Sliding surface is given as:

$$\sigma(X,t) = CE - W(t), \quad (19)$$

$C = [C_1 \ C_2]$ is normal constant matrix.

$$W(t) = CP(t), \quad (20)$$

$$P(t) = [p(t)^T \ \dot{p}(t)^T]^T, \ p(t) = [p_1(t) \ p_2(t) \ p_3(t)]^T.$$

and satisfy the following conditions: for a constant $T > 0$, $p_i(t)$ is bounded in $[0,T]$, and

$$p_i(0) = e_i(0), \dot{p}_i(0) = \dot{e}_i(0), p_i^{(2)}(0) = e_i^{(2)}(0), i = 1,2,3,$$

$$p_i(t) = \begin{cases} \sum_{k=0}^{n} \frac{1}{k!} e_i(0)^{(k)} t^k + \sum_{j=0}^{n} \left(\sum_{l=0}^{n} \frac{a_{jl}}{T^{j-l+n+1}} e_i(0)^{(l)} \right) \cdot t^{j+n+1}, 0 \le t \le T \\ 0, t > T \end{cases} \quad (21)$$

the order of this system $n = 2$.

The following coefficients are obtained by solving the equation:

$$\begin{cases} a_{00} = -10 \\ a_{10} = 15 \\ a_{20} = -6 \end{cases}, \quad \begin{cases} a_{01} = -6 \\ a_{11} = 8 \\ a_{21} = -3 \end{cases}, \quad \begin{cases} a_{02} = -1.5 \\ a_{12} = 1.5 \\ a_{22} = -0.5 \end{cases}.$$

Sliding surface is finally gained.

The terminal sliding mode controller is deduced by Lyapunov function as follows:

$$\dot{\sigma}(X,t) = CE - C\dot{P}(t) =$$

$$C\left[\dot{e}^T \ddot{e}^T \right]^T - C\left[\dot{p}^T(t) \ddot{p}^T(t) \right]^T =$$

$$C_2\left[\ddot{e} - \ddot{p}(t) \right] + \sum_{k=1}^{n-1} C_k\left[e^{(k)} - p^{(k)}(t) \right] =$$

$$C_2\left[\ddot{x}_1 - \ddot{x}_{1d} - \ddot{p}(t) \right] + C_1\left[\dot{e} - \dot{p}(t) \right] =$$

$$C_2\left[\dot{x}_2 - \ddot{x}_{1d} - \ddot{p}(t) \right] + C_1\left[\dot{e} - \dot{p}(t) \right],$$

$$\dot{\sigma}(X,t) = C_2[f(X,t) + \Delta f(X,t) + b(X,t)u +$$

$$d(t) - \ddot{x}_{1d} - \ddot{p}(t)] + C_1\left[\dot{e} - \dot{p}(t) \right] \qquad .$$

Lyapunov function is:

$$\dot{V} = \sigma^T \dot{\sigma} =$$

$$\begin{bmatrix} \sigma^T C_2\{f(X,t) - \ddot{x}_{1d} - \ddot{p}(t) + C_2^{-1}C_1[\dot{e} - \dot{p}(t)]\} \\ +\sigma^T C_2 b(X,t)u + \sigma^T C_2\left[\Delta f(X,t) + d(t)\right] \end{bmatrix} \le$$

$$\begin{bmatrix} \sigma^T C_2\{f(X,t) - \ddot{x}_{1d} - \ddot{p}(t) + C_2^{-1}C_1[\dot{e} - \dot{p}(t)]\} \\ +\sigma^T C_2 b(X,t)u + \| \sigma^T C_2 \| \cdot \| \Delta f(X,t) + d(t) \| \end{bmatrix}.$$

The final terminal controller is:

$$u(t) = \begin{cases} -b(X,t)^{-1}\{f(X,t) - \ddot{x}_{1d} - \ddot{p}(t) \\ +C_2^{-1}C_1(\dot{e} - \dot{p}(t))\} \\ -b(X,t)^{-1} \dfrac{C_2^T \sigma}{\|C_2^T \sigma\|}\{F(X,t) + D(t) + K\} \end{cases}. \quad (22)$$

K is positive constant.

$$\dot{V} \le \begin{cases} \| C_2^T \sigma \| \cdot \{\| \Delta f(X,t) + d(t) \| - [F(X,t) + D(t)]\} \\ -K \| C_2^T \sigma \| \end{cases} =$$

$$\| C_2^T \sigma \| \{[\| \Delta f(X,t) \| - F(X,t)] + [\| d(t) \| - D(t)]\} \qquad .$$

$$-K \| C_2^T \sigma \| \le -K \| C_2^T \sigma \| < 0$$

Based on the suppose and sliding surface Equation (19),

$$\sigma(X,0) = CE(0) - W(0) = CE(0) - P(0)$$

$$= C\left\{ \left[e(0)^T \dot{e}(0)^T \right]^T - \left[p(0)^T \dot{p}(0)^T \right]^T \right\} = 0 \quad.$$

The initial state of the system reach sliding surface, and global stability is guaranteed since the reach phase is removed. Therefore, it makes the systems stability and strong global robustness.

3.3 ADAPTIVE TERMINAL SLIDING MODE CONTROL WITH COMPENSATION TERM

The uncertainties such as small body rotating, solar radiation pressure and solar gravitational are more complex, so it is hard to obtain the boundary. The efficiency is not well if parameters in the control law are too large, on the contrary, the sliding mode is not conformed to exit if they are too small. Controller with the simple parameter adaptive law is effective for estimating the unknown parameters. This section presents a novel adaptive Terminal Sliding Mode Control law combined with variable structure and adaptive control for probe descent near the small body when the boundary of the uncertainties and external disturbances are unknown.

Suppose: The boundary of uncertainty $\Delta f(X,t)$ and external disturbance $d(t)$ are unknown, and satisfy the following inequality:

$$\| \Delta f(X,t) + d(t) \| \le \delta_0 + \delta_1 \| X \|, \quad (23)$$

δ_0 and δ_1 are unknown nonnegative constants.

The adaptive laws are proposed as follows to estimate the uncertainty and disturbance:

$$\dot{\hat{\delta}}_0(t,X) = \rho_0^{-1} \| C_n^T \sigma \|$$

$$\dot{\hat{\delta}}_1(t,X) = \rho_1^{-1} \| C_n^T \sigma \| \cdot \| X \| \qquad , \quad (24)$$

$$\hat{\delta}_0(t,X) = \bar{\delta}_0(t,X) - \delta_0, \hat{\delta}_1(t,X) = \bar{\delta}_1(t,X) - \delta_1 \qquad ,$$

$\bar{\delta}_0(t,X)$ and $\bar{\delta}_1(t,X)$ are the adaptive estimation of unknown parameters δ_0 and δ_1, ρ_0 and ρ_1 are positive adaptive gain. δ_0 and δ_1 are constants, so the Equation (24) can be rewritten as follow:

$$\dot{\bar{\delta}}_0(t,X) = \rho_0^{-1} \| C_n^T \sigma \|$$

$$\dot{\bar{\delta}}_1(t,X) = \rho_1^{-1} \| C_n^T \sigma \| \cdot \| X \| \qquad . \quad (25)$$

Considering Lyapunov function:

$$2V(\sigma,\hat{\delta}_0,\hat{\delta}_1) = \sigma^T\sigma + \rho_0\hat{\delta}_0^2 + \rho_1\hat{\delta}_1^2,$$

$$\dot{V}(\sigma,\hat{\delta}_0,\hat{\delta}_1) = \sigma^T\dot{\sigma} + \rho_0\hat{\delta}_0\dot{\hat{\delta}}_0 + \rho_1\hat{\delta}_1\dot{\hat{\delta}}_1 =$$

$$\left\{\sigma^T C_2 \begin{bmatrix} f(X,t) + \Delta f(X,t) + b(X,t)u + d(t) \\ -\ddot{x}_{1d} - \ddot{p}(t) \end{bmatrix} \atop +C_1[\dot{e} - \dot{p}(t)] + \rho_0\hat{\delta}_0\dot{\hat{\delta}}_0 + \rho_1\hat{\delta}_1\dot{\hat{\delta}}_1 \right\} \le \quad (26)$$

$$\left\{\sigma^T C_2[f(X,t) + b(X,t)u - \ddot{x}_{1d} - \ddot{p}(t)] \atop +C_1[\dot{e} - \dot{p}(t)] + \|C_n^T\sigma\|\cdot(\delta_0 + \delta_1\|X\|) \atop +\rho_0\hat{\delta}_0\dot{\hat{\delta}}_0 + \rho_1\hat{\delta}_1\dot{\hat{\delta}}_1 \right\}.$$

The final adaptive terminal controller is:

$$u(t) = -b(X,t)^{-1}\left\{ f(X,t) - \ddot{x}_{1d} - \ddot{p}(t) \atop +C_2^{-1}C_1(\dot{e} - \dot{p}(t)) \right\} -$$

$$b(X,t)^{-1}\frac{C_2^T\sigma}{\|C_2^T\sigma\|}\{(\bar{\delta}_0 + \bar{\delta}_1\|X\|) + K\}, \quad (27)$$

K is positive constant.

$$\dot{V}(\sigma,\hat{\delta}_0,\hat{\delta}_1) \le \left\{ \|C_2^T\sigma\|\cdot\{(\bar{\delta}_0 + \bar{\delta}_1\|X\|) + K\} \atop -\|C_2^T\sigma\|(\delta_0 + \delta_1\|X\|) \atop +\rho_0\hat{\delta}_0\dot{\hat{\delta}}_0 + \rho_1\hat{\delta}_1\dot{\hat{\delta}}_1 \right\} =$$

$$\left\{ -K\|C_2^T\sigma\| + \hat{\delta}_0(\rho_0\dot{\hat{\delta}}_0 - \|C_2^T\sigma\|) \atop +\hat{\delta}_1(\rho_1\dot{\hat{\delta}}_1 - \|C_2^T\sigma\|\cdot\|X\|) \right\} \le \quad (28)$$

$$-K\|C_2^T\sigma\| \le 0,$$

Based on Equations (26) and (28):

$$\dot{V}(\sigma,\hat{\delta}_0,\hat{\delta}_1) = \sigma^T\dot{\sigma} + \rho_0\hat{\delta}_0\dot{\hat{\delta}}_0 + \rho_1\hat{\delta}_1\dot{\hat{\delta}} \le -K\|C_2^T\sigma\|$$

That is:

$$\sigma^T\dot{\sigma} + \rho_0\hat{\delta}_0\rho_0^{-1}\|C_2^T\sigma\| + \rho_1\hat{\delta}_1\rho_1^{-1}\|C_2^T\sigma\|\|X\| \le -K\|C_2^T\sigma\|,$$

$$-\sigma^T\dot{\sigma} \ge \hat{\delta}_0\|C_2^T\sigma\| + \hat{\delta}_1\|C_2^T\sigma\|\cdot\|X\| + K\|C_2^T\sigma\| \ge 0$$

$$\|\sigma\|\cdot\|\dot{\sigma}\| \ge \hat{\delta}_0\|C_2^T\sigma\| + \hat{\delta}_1\|C_2^T\sigma\|\cdot\|X\| + K\|C_2^T\sigma\|$$

$$\|\dot{\sigma}\| \ge \frac{\hat{\delta}_0\|C_2^T\sigma\| + \hat{\delta}_1\|C_2^T\sigma\|\cdot\|X\| + K\|C_2^T\sigma\|}{\|\sigma\|}(\|\sigma\| \ne 0)$$

without loss of generality, let $C_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and:

$$\|\dot{\sigma}\| = \frac{\hat{\delta}_0\|\sigma\| + \hat{\delta}_1\|\sigma\|\cdot\|X\| + K\|\sigma\|}{\|\sigma\|}. \quad (29)$$

$$= \hat{\delta}_0 + \hat{\delta}_1\cdot\|X\| + K > K(\|\sigma\| \ne 0)$$

So the sliding mode σ convergences to zero in the finite time.

Meanwhile in order to reducing buffeting, $C_2^T\sigma/\|C_2^T\sigma\|$ is replaced by continuous function vector S_ε.

$$S_\delta = \frac{C_2^T\sigma}{\|C_2^T\sigma\| + \varepsilon}, \quad \varepsilon = \varepsilon_0 + \varepsilon_1\|e\|, \quad \varepsilon_0 \text{ and } \varepsilon_1 \text{ are two}$$

positive constant.

4 Simulation and analysis

The initial conditions and other simulation parameters of the target small body are listed in Table 1 referring to references [17, 18].

The trajectory of z axis is planning as Polynomial method mentioned in section 3.1 in order to landing on the surface of small body safely. The position curve $z_c(t)$ and velocity curve $\dot{z}_c(t)$ of z axis are as follows if the above initial condition and descent time $t_{go} = 4000s$ are given. Based on the Equations (13) - (16), z-axis is planed Nominal Polynomial trajectory.

$$z_c(t) = \left\{ z_0 + \dot{z}_0 t + (3z_n - 3z_0 - 2\dot{z}_0 t_{go})(t/t_{go})^2 \atop +(2z_0 + \dot{z}_0 t_{go} - 2z_n)(t/t_{go})^3 \right\}, \quad (30)$$

$$\dot{z}_c(t) = \left\{ \dot{z}_0 + (6z_n - 6z_0 - 4\dot{z}_0 t_{go})t/t_{go}^2 \atop +(6z_0 + 3\dot{z}_0 t_{go} - 6z_n)t^2/t_{go}^3 \right\}. \quad (31)$$

z_0 and $\dot{z}_0$ are the initial position and velocity of z axis, z_n and $\dot{z}_n$ is landing position and velocity of z-axis. The landing position and velocity of x-axis, y-axis and z-axis is desired as shown in table 1 in the finite time, and x-, y-axes are expected to zero but not the nominal polynomial trajectory.

TABLE 1 Parameters and Simulation value

Parameters	Simulation value
GM(m³/s²)	4.842×10⁵
Spin period(h)	10.54
Reference radius R0(m)	1138.5
C20	-0.003
C22	0.0396
Initial position	[350,300,9000]
Landing site	[0,0,7000]
Initial velocity	[-1.2,0.2,-1.0]
Landing velocity	[0,0,0]

473

Equation (18) and ATMSC with compensation term Equation (27) described above are incorporated into the simulation software that is used to generate time history of the probes dynamical state during the probe descent phase near rotating small body. The simulation results are shown in Figures 2-13 and all data are described in the Body-Fixed Coordinate *oxyz*.

The controller parameters are $K = 10$, $T = 10$, $\varepsilon_0 = 0.03$, $\varepsilon_1 = 1$, $\rho_0 = 1$, $\rho_1 = 1$:

$$C = [C_1 \quad C_2] = \begin{bmatrix} 1 & 0 & 0 & 4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 4 & 0 \\ 0 & 0 & 1 & 0 & 0 & 4 \end{bmatrix}.$$
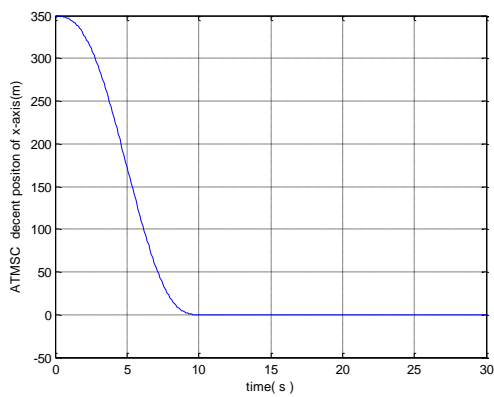
FIGURE 2 ATSMC descent altitude time history of *x*-axis
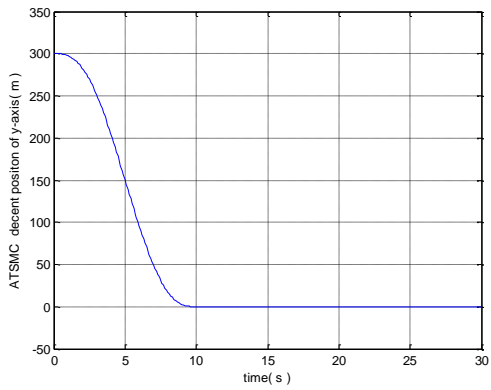
FIGURE 3 ATSMC descent altitude time history of *y*-axis
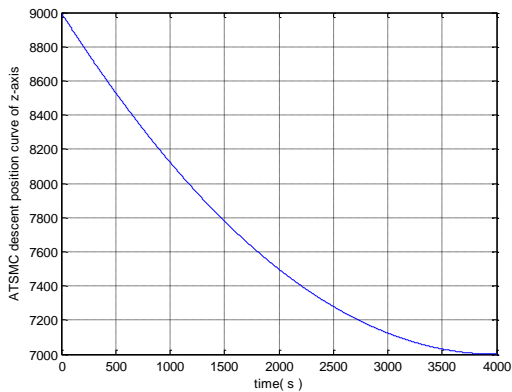
FIGURE 4 ATSMC descent altitude time history of *z*-axis
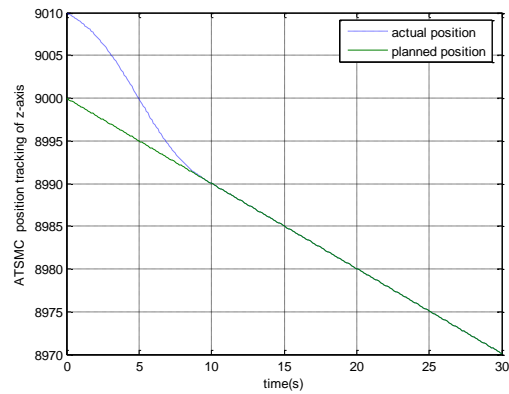
FIGURE 5 ATSMC descent position tracking of *z*-axis
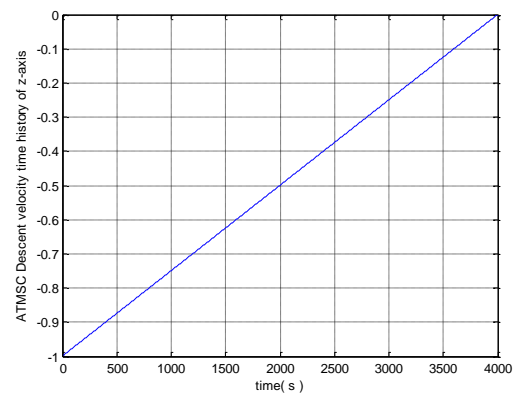
FIGURE 6 ATSMC descent velocity time history of *z*-axis
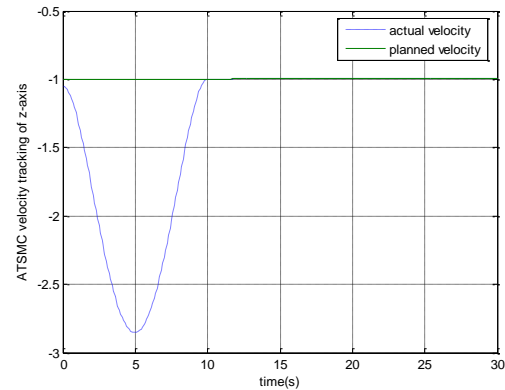
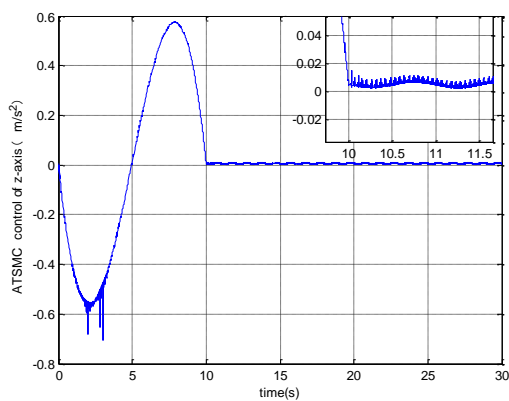FIGURE 7 ATSMC descent velocity tracking of *z*-axis

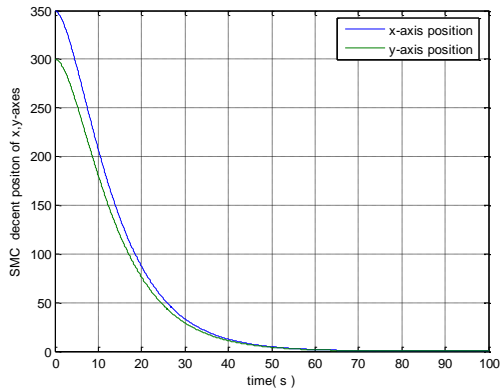FIGURE 8 Control acceleration of *z*-axis

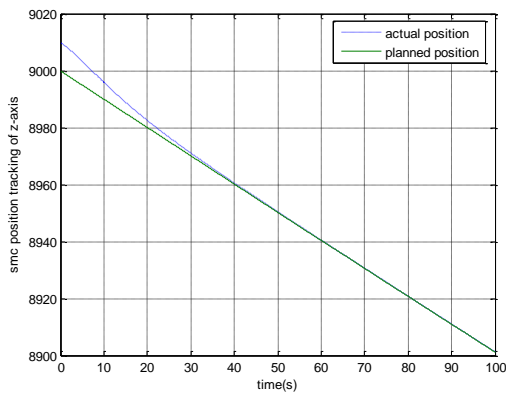FIGURE 9 SMC descent altitude time history of *x*, *y*-axes

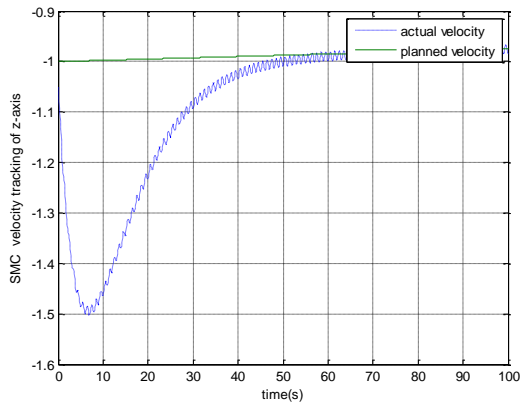FIGURE 10 SMC descent position tracking of *z*-axis

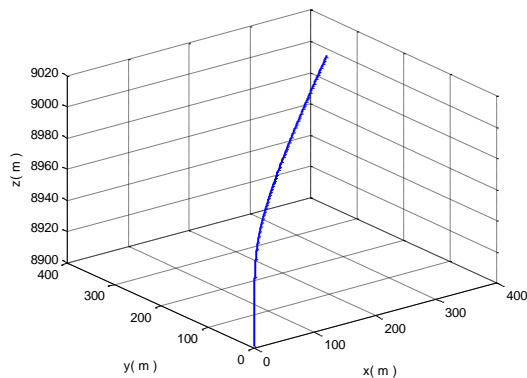FIGURE 11 SMC descent velocity tracking of *z*-axis

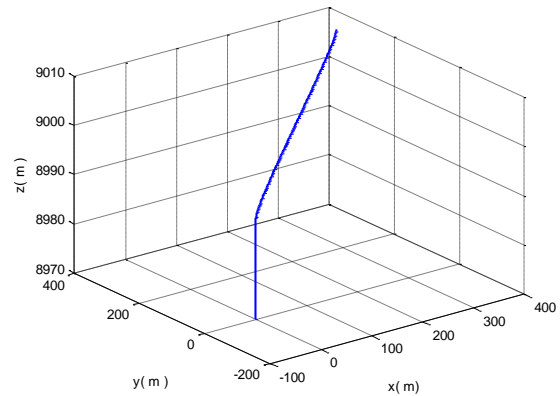

FIGURE 12 SMC three-dimensional trace

FIGURE 13 ATSMC three-dimensional trace

Figures 2-4 give the descent position simulation time history of *x*-, *y*-, and *z*-axes respectively under the controller Equation (27) and here all data are described in the Body-Fixed Coordinate system. Figure 6 gives the velocity simulation time history of *z*-axis under controller Equation (27). The descent positions of *x*- and *y*- axes approach zeros quickly, meanwhile descent position and descent velocity of *z*-axis conform to the planed Nominal Polynomial trajectory Equations (23) and (24). Figures 5 and 7 give the descent position and velocity tracking curves of *z*-axis. These results show that positions and velocity of *x*-, *y*- and *z*-axes can reach desired site in the finite time about 10s. Control force in Figure 8 can satisfied engineering requirements. Figures 9, 10 and 11 give the descent position and velocity time simulation curve of *x*-, *y*-, and *z*-axes based on normal Sliding Mode Control. The tracking time is about 50s and the curves are affected by perturbation obviously. We can see the guidance and control algorithm based on Adaptive Terminal Sliding Mode Control can fast and accurately track the planned trajectory in the finite time and is more robust to parameter uncertainty, feedback state error and external disturbances than the normal Sliding Mode Control. The performance of tracking the desired trajectory fast is also verified from three-dimensional curves in Figure 12 and 13.

5 Conclusions

The present paper examines the three dimensional orbital dynamics and control aspects of the probe descent near rotating small bodies. The relative kinetics equations for the system are derived using Newton's second law in the Body-Fixed Coordinate system with the assumption that the small body is a homogeneous tri-axial ellipsoid. The nominal polynomial trajectory is employed for *z*-axis taking account of the fuel suboptimal. Terminal Sliding Mode Control law with compensation term is derived and the efficacy of the controller performance is tested via numerical simulation. The Adaptive Terminal Sliding Mode Control law with compensation term is then derived in the case the boundary of uncertainties and external

disturbances are unknown. Results of the numerical simulation match with closed-form solutions and demonstrate the proposed algorithm can ensure fast and accurately respond in conditions of parameter uncertainty, feedback state error and external disturbances compared with SMC, and the system can ensure the descent accuracy requirement and a safe landing velocity.

6 Further Work

The gravitational potential function adopted in the present paper is the second order harmonic series expansion method which may generate inaccuracy when the probe closing to the small body. Segmentation method is intended to adopted for calculating the gravitational potential function when the probe landing on the surface of small body. Harmonic series expansion method is adopted when the probe is far from the small body and polyhedral approach is adopted when the distance is close.

Nominal polynomial trajectory in the present paper has limitations only considering fuel suboptimal. The future work is focused on descent trajectory planning taking fully account into the fuel and irregular gravitation etc.

References

[1] Kowal C T 1996 Asteroids Their nature and utilization *Wiley*, Chichester UK

[2] Veverka J, Farquhar B, Robinson M, Thomas P, Murchie S, Harch A, Antreasian PG, Chesley SR, Miller JK, Owen WM Jr, Williams BG, Yeomans D, Dunham D, Heyler G, Holdridge M, Nelson RL, Whittenburg KE, Ray JC, Carcich B, Cheng A, Chapman C, Bell JF 3rd, Bell M, Bussey B, Clark B, Domingue D, Gaffey MJ, Hawkins E, Izenberg N, Joseph J, Kirk R, Lucey P, Malin M, McFadden L, Merline WJ, Peterson C, Prockter L, Warren J, Wellnitz D 2001 *Nature* **413**(6854) 390-3

[3] Holdridge M E 2002 *Johns Hopkins APL technical digest* **23**(1) 58-70

[4] David W D, Robert W F, Farquhar R W, McAdams J V 2002 *Icarus* **159**(2) 433-8

[5] Kubota T, Hashimoto T, Sawai S, Kawaguchi J I, Ninomiya K, Uo M, Baba K 2003 *Acta Astronautica* **52**(2) 125-131

[6] Kubota T, Otsuki M, Hashimoto T, Bando N, Yano H, Uo M, Kawaguchi J I 2006 *Proceedings of AIAA/AAS Astrodynamics Specialist Conference and Exhibit 2006* AIAA-2006-6539

[7] Varatharajoo R, Kahle R 2005 *Aircraft Engineering and Aerospace Technology* **77**(2) 131-41

[8] Li S, Cui H T, Cui P Y 2005 *Aircraft Engineering and Aerospace Technology* **77**(4) 317-23

[9] Cui H T, Shi X Y, Cui P Y 2002 *Chinese Journal of Space Science* **22**(3) 256-60

[10] Colin R M, Gianmarco R 1996 *Journal of Guidance, Control and Dynamics* **19**(3) 740-2

[11] Carson J M, Açıkmese B 2006 A model predictive control technique with guaranteed resolvability and required thruster silent times for small-body proximity operations *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit* Keystone CO AIAA-2006-6780

[12] Bellerose J, Girard A, Scheeres D J 2009 Optimization and Cooperative Control Strategies, *Springer Berlin Heidelberg* 135-150

[13] Liu J K 2005 *MATLAB Simulation for sliding mode control* Tsinghua Press: Beijing *(in Chinese)*

[14] Hu J B, Shi M H 2005 *Control Theory &Applications* **22**(3) 540-8 *(in Chinese)*

[15] Zhou J Y, Zhou D 2007 *Journal of Astronautics* **28**(6) 1462-6 *(in Chinese)*

[16] Werner R A, Scheeres D J 1996 *Celestial Mechanics and Dynamical Astronomy*, **65**(3) 313-44

[17] Zhang Z X, Wu W D, Li L T, Huang X Y, Cui H T, Li S, Cui P Y 2012 *Journal of the Franklin Institute* **349**(2) 493-509

[18] Huang X Y, Cui H T 2004 *Acta Astronautica* **54**(10) 763-71

Authors

Chunhui Liang, born in January, 1977, Fuxin City, Liaoning Province, China

Current position, grades: PhD student at Jilin University, Lecturer at Changchun Institute of Technology.
University studies: Control theory and control engineering.
Scientific interest: Intelligent control, guidance and control.
Publications: more than 10 academic papers.

Yuanchun Li, born in April, 1962, Jilin City, Jilin Province, China

Current position, grades: Ph D, Professor in department of control theory and control engineering, communication engineering institute, JiLin University, ChangChun City, China. Director at College of Electrical Engineering, Changchun University of Technology.
University studies: PhD degree at Harbin Institute of Technology.
Scientific interest: robot control, vehicle control, intelligent control, guidance and control.
Publications: more than 10 academic papers.

Research on modelling of intake tower in three-dimension CAD software and simulation analysis in FE software

Hongyang Zhang*

North China University of Water Resources and Electric Power, Zhengzhou, China

Abstract

ANSYS is one of the most influential finite element analysis software in the world because of its very powerful calculation and analysis ability, but its pre-process function is weak relatively. SolidWorks is the three-dimension parametric feature modelling software of 100% feature modelling and 100% parameterization, which provides product-level automated design tools. In this paper, combining with the intake tower, it discusses the method of modelling in three-dimension CAD software SolidWorks and the interface processing between SolidWorks and the ANSYS code, which decreases the difficulty in modelling complicated models in ANSYS. In view of the function of the birth-death element and secondary development with APDL (ANSYS parametric design language), simulation analyses of thermal field and stress during the construction and impounding periods were conveniently conducted.

Keywords: computer modelling, SolidWorks, interface processing, birth-death element, APDL, simulation analysis, ANSYS

1 Introduction

Recently, mass concrete is widely applied to hydraulic engineering. The internal temperature of concrete rises due to cement hydration heat, and thus generates thermal stress of the concrete structure. Excessive stress may cause concrete cracks, which affects safety of the concrete structure. It is necessary to analyse the thermal field and thermal stress of important mass concrete structures with both routine methods and the finite element method (FEM). Some researchers have done a large amount of simulation analyses using FE software [1-6], but difficulties in these methods remain. There are two main difficulties:

(i) Modelling of mass concrete structures in FEM software are difficult because of their complexity [7, 8].

(ii) The construction processes and boundary conditions of concrete structure are complex, so complete simulation is difficult with FEM software [9, 10].

How to solve these two difficulties is of concern to engineers and researchers.

SolidWorks is a CAD/CAE/CAM/PDM desktop system, and the first three-dimension mechanical CAD software in Windows developed by the SolidWorks company. It provides product-level automated design tools [11]. Since its introduction in 1995, SolidWorks has become a favourite design tool for many of today's engineers, mechanical designers, and industrial designers. In part because of its easy-to-learn graphical user interface and powerful set of tools. SolidWorks is used by many top companies worldwide to design, engineer, and document their products in a variety of fields. At the core of SolidWorks is the ability to create parametric three-dimension solid geometry that is then used to create drawings, manufacturing instructions, instruction manuals, animations, full-colour renderings, and other types of documentation. Regardless of the complexity of the item being created, the creation process is easy and follows the same basic steps. Firstly, a sketch is created that is turned into a base feature. The base feature is then further refined by adding features that add or remove material from the base feature. Individual part models can then be used to build assemblies that represent the final design. After creating the three-dimension part or assembly models, drawings are made to document the design and manufacturing process. SolidWorks is the three-dimension parametric feature modelling software of 100% feature modelling and 100% parameterization, which has the following outstanding characteristics.

(i) Characteristics of the administrator functions.

(ii) Full-related data management of the zero part design, assembly design and two-dimension drawings, which are interrelated.

(iii) With tightly interface of a number of CAM, CAE software.

At present, modelling by SolidWorks is mainly applied to railway, aerospace, machinery manufacturing, national defence industry, electronics, shipbuilding and other fields [12-15], and is rarely applied to hydraulic engineering.

ANSYS is a type of large universal finite element software that has a powerful ability to calculate and analyse aspects of structure, thermal properties, fluid, electromagnetic, acoustics and so on, which has been widely used in civil engineering, traffic, water conservancy, railway, petroleum chemical industry, aerospace, machinery manufacturing, national defence,

---
* Corresponding author e-mail: 1694440@qq.com

electronics, shipbuilding, biological medicine, geological mining, household appliances and other general industrial and scientific research [16]. Since developed in 1970 by John Dr Swanson at the university of Pittsburgh, ANSYS has occupied a pivotal position in the FES field and been widely accepted by the industrial areas, which is recognized as the standard analysis software of more than 20 professional and technical associations. But the pre-process function of ANSYS is weak relatively. The modelling ability in ANSYS is inferior and the complex model building is very complex by ANSYS. So the reprocessing wastes most time and affects work efficiency seriously.

In the paper, the structure of intake tower was modelled in the three-dimension CAD software SolidWorks and imported into ANSYS with an interface tool. Then, the simulation analysis during construction and impounding periods was conducted by the APDL program in ANSYS.

2 Modelling in SolidWorks and interface processing between SolidWorks and ANSYS

2.1 MODELLING IN SOLIDWORKS

The process of modelling is undertaken from the inside to the outside because of the complexity of internal structure of the intake tower.

(i) The base plate and the left side wall modelling. Firstly, selecting sketch map and using sketch rendering tools to draw L-shaped cross-section of base plate and side wall. Secondly, entering the values of stretching length and stretching the cross-section by stretching convex body command, and the three-dimension model of base plate and the left side wall is obtained, which is shown in Figure 1.

(ii) Internal structure modelling. Firstly, selecting sketch map and drawing the two-dimension plan for stretching. Secondly, entering the values of stretching length and stretching the plan by stretching convex body command, and the three-dimension model internal structure is obtained. The internal structure model is shown in Figure 2.
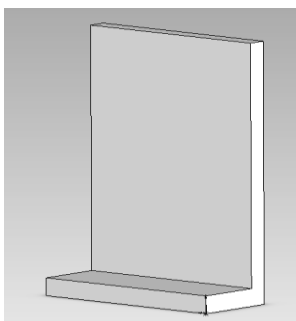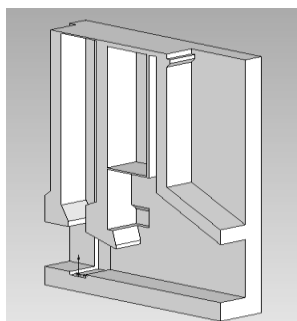

FIGURE 1 3D L-shaped model

FIGURE 2 Internal structure model

(iii) Optimal model. Modelling the right side wall and modifying the model by stretching removal command. The optimal model is shown in Figure 3. In addition, SolidWorks provides the command of insert → characteristics → split for model splitting, by which the simulation analyses for the construction impounding of the mass structure become convenient.

(iv) Examining the section of model from different perspectives. In SolidWorks, we can easily check the front and back, up and down, left and right sides of model, and even different location of the section, which helps us conveniently check whether the model is optimal. The cross section of model is shown in Figure 4.
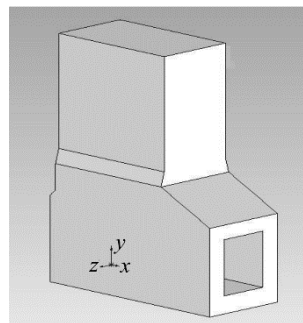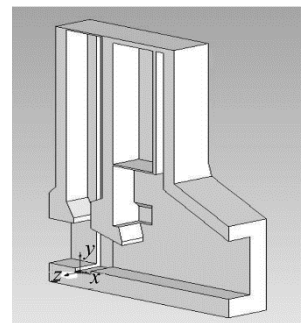

FIGURE 3 Integrated model   FIGURE 4 Cross section

2.2 INTERFACES BETWEEN SOLIDWORKS AND ANSYS

ANSYS provides some interface tools for CAD software, which imports the CAD model conveniently and reduces the difficulties of model processing. The interface tools are given in Table 1.

TABLE 1 CAD software packages and preferred interface tools

CAD software package	File type	Interface tool
AutoCAD	*.sat	Interface tool for SAT
Pro/ENGINEER	*.prt	Interface tool for Pro/ENGINEER
SolidWorks	*.x_t	Interface tool for Parasolid

The model needs to save as type Parasolid (*.x_t) to import into ANSYS correctly, concrete steps are as follows, choose "file → save as", it is requested to write the model's name and choose the saving type as Parasolid (*.x_t) in the ejecting dialog box, then choose the saving folder and save. In ANSYS, using the command "PARAIN, Name, Extension, Path, Entity, FMT, Scale" or choosing "File → Import → PARA..." in the GUI interface. There are two means of importing, and the differences of whether selecting "Allow Defeaturing" or not are given in Figures 5 and 6. Both of models can be pre-processed in ANSYS in practice.
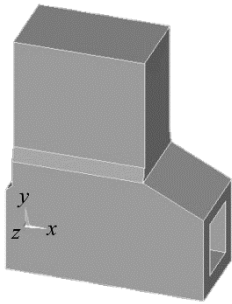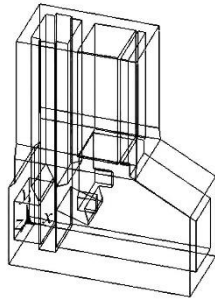
478

FIGURE 5 With defeaturing　　　FIGURE 6 Without defeaturing

3 Analysis of thermal field of intake tower

There are 5 products of ANSYS to thermal analysis as shown in Figures 7, which can be classified into ANSYS/ Multiphysics, ANSYS/Mechanical, ANSYS/Thermal, ANSYS/FLOTRAN and ANSYS/ED.
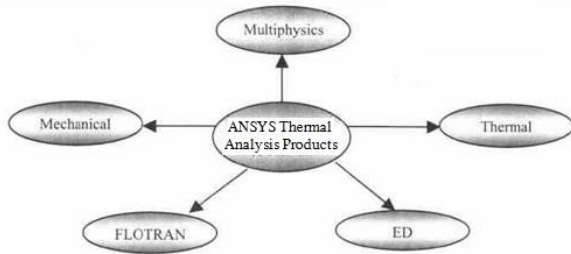
FIGURE 7 ANSYS thermal analysis products

The heat balance equilibrium equation of thermal analysis in ANSYS is based on the principle of energy. In calculating, the temperature of each node is calculated, and then other thermal physical parameters are exported. and export the finite element method to calculate. It can figure out the problems of heat conduction, convection and radiation. In addition, the problems of phase transition, inner heat source and the thermal contact resistance also can be analysed [17-19].

The thermal analysis in ANSYS can be classified into steady heat transfer analysis and transient heat transfer analysis. The steady heat transfer is that the thermal field does not change with time, and the transient heat transfer is that the thermal field changes significantly with time. The temperature analysis of the intake tower during the construction and impounding periods involves aspects of the thermal field and thermal stress. The calculation must deal with the problems of simulation of layered construction, dynamic boundary conditions, hydration heat, dynamic elasticity modulus, autogenously volume deformation of concrete and thermal creep stress, which are difficult to simulate directly in ANSYS. APDL is a scripting language based on the style of parametric variables. It is used to reduce a large amount of repetitive work in analysis [20-22]. This study carried out a simulation analysis of the thermal field considering nearly all conditions of construction, using the birth-death element and programming with APDL.

Killing the elements is not removing the element from the model, but reducing its stiffness by multiplying a small factor with its stiffness matrix. At the same time, the element loads, strains, mass, damp, specific heat and other similar parameters of the dead element are set to zero, which are not included in the results. As the same process as above, the activated element is not added to the model, but to re-activate it, then the stiffness, mass and other parameters return to the original values. All elements must be generated in PREP7, including to the dead elements which will be activated later, because the elements cannot be generated in equation solver. In analysis, all elements are killed at the beginning and the elements which are needed to analyse are re-activated later.

The commands of setting birth-death element by APDL are written as:

ESEL, … ! Selecting the elements that will be killed
EKILL, … ! Killing the elements
ESEL,… ! Selecting the elements that will be activated
EALIVE, … ! Activating the elements

It can be used the loop commands to realize the layered construction of concrete, which solves the problem of simulating difficultly in GUI interface of ANSYS. The loop commands are written as:

*DO, Par, IVAL, FVAL, INC
Commands section
*ENDDO

The written commands can be saved in text file. Using the command "/INPUT, 'file name', 'txt', 'saving path' " or choosing "File read input from..." in the GUI interface, and importing the text file with commands, which realize the simulation of construction progress conveniently.

The life and death of element functions are realized by modifying the stiffness. When the element was "killed", the stiffness matrix of which is not removed, but its values reduced to a minimum. The stiffness of killed elements multiplied by a tiny reduction factor (default is 1e-6), and the values of the stiffness is not zero to prevent matrix singularity.

3.1 ANALYSIS OF THERMAL FIELD OF INTAKE TOWER

3.1.1 Unsteady thermal field analysis

Many factors such as the effect of hydration heat of cement, air temperature and water temperature cause the temperature of concrete changes. This is a heat conduction problem of internal heat sources in the area. The unsteady thermal field is written as [23]:

$$\frac{\partial T}{\partial \tau} = \frac{\lambda}{c\rho}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + \frac{\partial \theta}{\partial \tau}, \tag{1}$$

where τ is the age of concrete, c is the specific heat of concrete, ρ is the density of concrete, λ is the thermal

conductivity of concrete, and θ is the adiabatic temperature rise of concrete.

For the three-dimension unsteady thermal field, the functional form $I^e(T)$ is:

$$I^e(T) = \iiint_{\Delta R} \left\{ \frac{1}{2}\alpha \left[\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2 \right] + \right.$$

$$\left. \iiint_{\Delta R} \left(\frac{\partial T}{\partial \tau} - \frac{\partial \theta}{\partial \tau}\right) T \right\} dxdydz + \iint_{\Delta D} \bar{\beta}\left(\frac{1}{2}T^2 - T_a T\right) ds, \quad (2)$$

where ΔR is a subfield of unit e, the thermal diffusivity $\alpha = \dfrac{\lambda}{c\rho}$, $\bar{\beta} = \dfrac{\beta}{c\rho}$, β is the exothermic coefficient, ΔD is the area on surface D, which is only in boundary units, and T_a is the air temperature.

3.1.2 Initial and boundary conditions

In analysis, the calculated initial temperature of concrete is 10°C.

The index formula of hydration heat of cement is written as [24, 25],

$$Q(t) = 71610 \left[1 - \exp(-0.36t) \right], \quad (3)$$

where t is the pouring time, and Q is the hydration heat. The relation between Q and θ is written as:

$$\frac{\partial \theta}{\partial \tau} = \frac{Q}{c\rho}. \quad$$

The boundary conditions involve the laws of interaction between concrete and the surrounding medium. When concrete is exposed to the water, the boundary condition is:

$$T(\tau) = f(\tau). \quad (5)$$

When concrete is exposed to the air, the boundary condition is:

$$-\lambda\left(\frac{\partial T}{\partial n}\right) = \beta(T - T_a), \quad (6)$$

where n is the normal direction. Both T_a and β are constants or variables.

The steel formworks and straws are used as the insulation materials during the maintenance period, and the exothermic coefficients of which are 45 kJ/(m²h°C) and 10 kJ/(m²h°C), respectively.

The air temperature variation formula is written as:

$$T = 26.1 - 25.1\cos\left[\frac{\pi}{284}(t - 79)\right]. \quad (7)$$

The air temperature formula during impounding period is written as,

$$T = 15.8 + 11.9 \times \cos[\pi \times (t - 281)/180]. \quad (8)$$

After impounding finished, the following formulas are given to describe the water temperature variation of different height. Height from 89m to 107m, $T = 18°C$.

Height from 107m to 110m:

$$T = 18 + 2 \times \cos[\pi \times (t - 281)/180]. \quad (9)$$

Height from 110m to 112m:

$$T = 18 + 3 \times \cos[\pi \times (t - 281)/180]. \quad (10)$$

Height from 112m to 113m:

$$T = 18 + 5 \times \cos[\pi \times (t - 281)/180]. \quad (11)$$

3.2 THERMAL FIELD ANALYSIS IN ANSYS

Table 2 shows the construction scheme of layered construction. A layer is not poured until the former layer is poured. The pouring days in Table 2 are all the total days of construction for each layer.

TABLE 2 Construction scheme

Construction elevation (m)	Pouring day (d)	Construction elevation (m)	Pouring day (d)
86.5-89.0	1-25	103.0-108.0	122-152
89.0-95.0	26-50	108.0-114.0	153-179
95.0-96.8	51-89	114.0-120.0	180-201
96.8-103.0	90-121	120.0-121.0	202-221

In the analysis, the simulation of layered construction and impounding is settled conveniently by function of the birth-death element and secondary development with APDL (ANSYS parametric design language).

The coordinates and maximum temperatures of feature points in every layer are given in Table 3 and the temperature curves are shown in Figure 8.

TABLE 3 Coordinates and maximum temperature of feature points

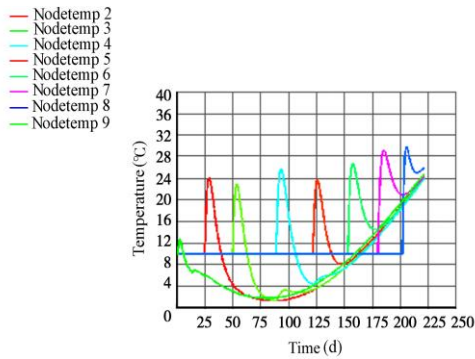| Feature point number	x	y	z	height	Maximum temperature (°C)
2	7.4	6.0	1.0	92.5	24.268
3	8.4	9.9	5.0	96.4	24.353
4	16.4	16.0	5.0	102.5	25.743
5	8.4	20.0	5.0	106.5	24.346
6	16.4	25.0	8.0	111.5	26.611
7	1.0	30.0	2.0	116.5	29.224
8	8.4	34.5	5.0	121.0	29.934

FIGURE 8 Maximum temperature curves

In Figure 8, the curves from Nodetemp 2 to Nodetemp 8 express temperature variation of feature points from 2 to 8, and the curve of Nodetemp 9 is the air temperature curve. It is shown that the maximum temperature occurs on the 3rd or 4th day after pouring and decreases with time. Feature point 4, the coordinates of which are (16.4, 16.0, 5.0), shows the maximum temperature difference of 23.534°C. Feature point 8 shows the maximum temperature rise during the construction period, and the maximum temperature of which is 29.934°C, occurring on the 206th day of the total construction period.

The impounding commenced immediately after the construction of the intake tower, which lasted for 16 days, and the calculating lasted for 170 days. In particular, the reservoir elevation started from 84.0m and finished at 108.0m, at the rate of approximately 1.5m per day. The feature points are selected in every layer above the base plate. The maximum temperatures and the temperature curves are given in Table 4 and Figure 9, respectively.

TABLE 4 Coordinates and maximum temperature of feature points

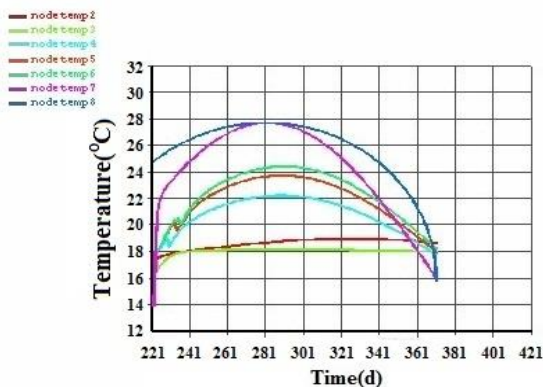| Feature point number	x	y	z	height	Maximum temperature (°C)
2	5.0	2.0	23.0	88.5	18.979
3	1.0	7.0	23.0	93.5	18.145
4	10.0	15.04	7.632	101.54	22.221
5	5.0	24.0	8.632	110.5	23.729
6	5.0	26.0	24.0	112.5	24.413
7	5.0	34.5	17.632	121.0	27.698
8	Air temperature variation curve				

FIGURE 9 Maximum temperature curves

Figure 9 shows that the maximum temperature of each layer occurs on the 60th day after impounding, and then the temperature decreases with time. In Figure 6, the numbers of feature points from 2 to 7 are corresponding to their maximum temperature curves from Nodetemp 2 to Nodetemp 7, and the curve of Nodetemp 8 is the air temperature curve. Feature point 6, the coordinates of which are (5.0, 26.0, 24.0), the maximum temperature of which underwater is 24.413 °C, occurring on the 285th day of the total construction and impounding periods.

4 Analysis of thermal stress of intake tower

Expansion or contraction of the structure occurs along with heating and cooling. The thermal stress occurs when the expansion or contraction is limited. In this paper, the temperature of nodes was applied to the structure as a body load after the analysis of the thermal field.

4.1 SELECTION OF CALCULATING PARAMETERS

The parameters of concrete are given in Table 5.

TABLE 5 Parameters of concrete

Material	Density (kg/m³)	Coefficient of linear expansion (1/°C)	Poisson ratio
Concrete	2 447.5	9×10^{-6}	0.167

The elasticity modulus is written as:

$$E_1 = 3.6 \times 10^{10} \left[1 - \exp\left(-0.40 t^{0.34}\right) \right]. \qquad (12)$$

The creep effect is considered in analysis of temperature stress, and the formula of the creep degree is written as,

$$C = \left[0.23\left(1 + 9.2t^{-0.45}\right)\left(1 - e^{t_1}\right) + 0.52\left(1 + 1.17t^{-0.45}\right) \times \right.$$
$$\left. \left(1 - e^{t_2}\right) \right] \times 10^{-10} \Big/ 3.60. \qquad (13)$$

The creep degree is influenced by the cement type, water-cement ratio and admixture.

where $t_1 = -0.3(t - 3)$ and $t_2 = -0.005(t - 3)$. When $t_1 < -80$, we consider $t_1 = -80$; and when $t_2 < -80$, we consider $t_2 = -80$.

Considering the creep degree, the formula of the elasticity modulus is adjusted to be:

$$E = E_1 / \left(1 + CE_1\right). \qquad (14)$$

4.2 THERMAL STRESS ANALYSIS IN ANSYS

ANSYS software provides the following two methods of thermal stress analysis:

(i) The indirect method. Thermal field analysis is constructed first, and then the node temperature is applied to the structure as a body load.

(ii) The direct method. The results of thermal field and thermal stress are gained by using coupling elements with both temperature and displacement degree of freedom.

In this paper, the thermal field and thermal stress analyses during the construction and impounding periods belongs to the conditions that temperature of nodes are unknown and thermal and structural coupling is unidirectional, so the first method which is the indirect method is selected in analysis.

The coordinates of feature points in thermal stress analysis were same as those in thermal field analysis. Table 6 shows the maximum thermal stress of each point. Feature point 9 is the point with the maximum thermal stress.

TABLE 6 Maximum thermal stress of feature points

Feature point number	Maximum thermal stress (MPa)	Feature point number	Maximum thermal stress (MPa)
2	0.25	6	0.17
3	0.26	7	0.13
4	0.38	8	0.14
5	0.37	9	1.68

The thermal stress curves of feature points are shown in Figure 10.
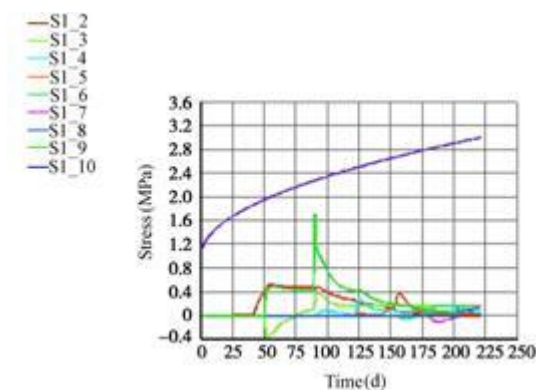

FIGURE 10 Maximum stress curves

Here the curves from S1_2 to S1_9 express the maximum stress variation of feature points from 2 to 9, and the S1_10 curve is the ultimate tensile stress of concrete. The formula of concrete's ultimate tensile stress is written as:

$$\sigma_t = 0.232 \times 10^6 \left\{ 33.5 \left[1 + 0.2 \times \ln\left(t/28 \right) \right] \right\}^{\frac{2}{3}}. \qquad (15)$$

The maximum thermal stress occurs on 90th day of the construction period on feature point 9, which is located at the interface between the third layer and the fourth layer, and the value of which is 1.68 MPa. It is known that the thermal stress increases with the temperature difference. Thus, it is postulated that the maximum thermal stress is caused by the instantaneous temperature difference between two layers in the pouring period. It is known that the maximum thermal stress of each point during the construction period is less than the ultimate tensile stress of concrete from Figure 10.

The maximum thermal stress of each point during impounding period is shown in Table 7. Feature point 6, the coordinates of which are (5.0, 26.0, 24.0), is the point with the maximum thermal stress.

TABLE 7 Maximum thermal stress of feature points

Feature point number	Maximum thermal stress (MPa)	Feature point number	Maximum thermal stress (MPa)
2	--	5	0.22
3	0.10	6	0.26
4	0.12	7	0.23

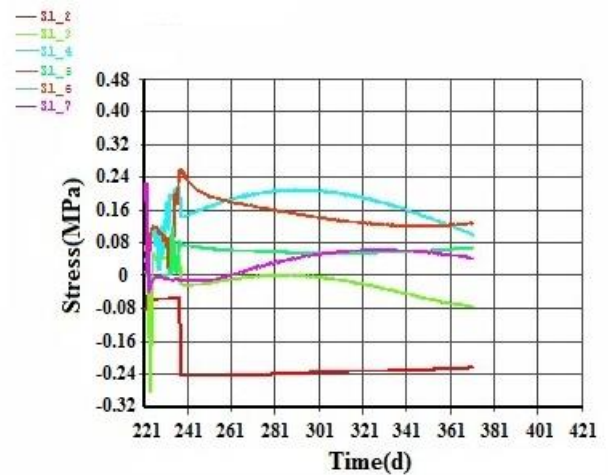The thermal stress curves of feature points are shown in Figure 11.

FIGURE 11 Maximum stress curves

Here the numbers of feature points from 2 to 7 are corresponding to their maximum stress curves from S1_2 to S1_7. The figures and table show that the maximum thermal stress of the intake tower is 0.26 MPa, occurring on the 18th day of the impounding period, and the maximum thermal stress of each point in the intake tower during the impounding period is less than the ultimate tensile stress of concrete.

5 Conclusions

(i) In this paper, the method of modelling in three-dimension CAD software SolidWorks and the interface processing between SolidWorks and the ANSYS code are discussed, which realizes an effective combination of the advantages of both SolidWorks and ANSYS.

(ii) The thermal field and thermal stress during the construction and impounding periods, considering multi factors such as layered construction and impounding, hydration heat, dynamic elasticity modulus, autogenous volume deformation and creep of concrete, are conveniently obtained by birth-death element and secondary development with APDL.

(iii) It is known that the temperature rises rapidly in the early stage of construction, and reaches a maximum value of 29.934°C on the 3rd or 4th day

after pouring. The maximum temperature during impounding period is 24.413°C, occurring on the early stage of impounding. The thermal stress increases with the temperature difference, and the maximum thermal stress occurs at the interface of new and old layers, which is caused by the instantaneous temperature difference. The values of maximum thermal stress of construction and impounding are 1.68 MPa and 0.26 MPa, which

are both less than the ultimate tensile stress of concrete.

Acknowledgments

References

[1] Ashida F, Tauchert T R 1998 An inverse problem for determination of transient surface temperature from piezoelectric sensor measurement *ASME Journal of Applied Mechanics* **65**(2) 367-73

[2] Kawaguchi T, Nakane S 1996 Investigations on determining thermal stress in massive concrete structures *ACI Materials Journal* **93**(1) 96-101

[3] Lin J Y, Cheng T F 1997 Numerical estimation of thermal conductivity from boundary temperature measurements *Numerical Heat Transfer* **32**(2) 187-203

[4] Wu Y, Luna R 2001 Numerical implementation of temperature and creep in mass concrete *Finite Elements in Analysis and Design* **37**(2) 97-106

[5] Wu J G, Li X J, He K F 2012 Simulation of Rolling Forming of Precision Profile Used for Piston Ring based on LS_DYNA *Journal of Computers* **7**(9) 2208-15

[6] Li X J, Wang K, Jiang L L 2012 Rotor Crack Detection by Using Multi-vibration Signal from the Basement *Journal of Software* **7**(5) 959-65

[7] Zhang J 2005 Interface design between AutoCAD and ANSYS *Chinese Quarterly of Mechanics* **26** 257-62 *(in Chinese)*

[8] Du P, Liu S X, Tan G Z, Liu K X 2012 Numerical analysis of mass concrete based on the theory of four dimensional thermal field *Journal of Liaoning Technical University (Natural Science)* **31**(4) 526-30 *(in Chinese)*

[9] Zhao Y W, Geng D X, Liu X M 2012 An Investigation on Electrorheological Fluid-assisted Polishing Process for Tungsten Carbide. *Journal of Computers* **7**(11) 2742-9

[10] Jin K K, Kook H K, Yang J K 2001 Thermal analysis of hydration heat in concrete structures with pipe-cooling system *Computers and Structures* **79**(2) 163-171

[11] Liu L J, Ren J P 2005 Application of the secondary development in SolidWorks *Mechanical Management and Development* **20** 74-5 *(in Chinese)*

[12] Prabhakar S, Henderson M R 1992 Automatic form-feature recognition using neural network-based techniques on boundary representation of solid models *Computer–Aided Design* **24**(7) 381-93

[13] Rajan G, Venkat D 1998 Intersecting Features Extraction Form two-dimension Orthographic Projection *Computer–Aided Design* **30**(11) 863-73

[14] Gao Y 2010 A probe into gear modeling based on SolidWorks *Journal of Yangzhou Polytechnic College* 14(**3**) 22-4 *(in Chinese)*

[15] Li X, Huang Z F, Li H T 2013 Geometric and Kinematics Modeling of Tele-operated Virtual Construction Robot *Journal of Software* **8**(10) 2517-21

[16] Gao P, Li P H, Bao K M 2012 Research on coupled heat and moisture transfer of mass concrete *Journal of China Three Gorges university* **4** 29-33

[17] Chen G R, Xu W T, Yang Y, Li K 2012 Computation Method for Thermal field of Mass Concrete Containing Cooling Water Pipes *Chinese Journal of Computational Physics* **29**(3) 411-6

[18] Li X F 2012 Nonlinear forecast for temperature of massive concrete *Concrete* **22** 127-9 *(in Chinese)*

[19] Chen S K, Zhang Y P, Guo L X, Wang H B, Xie Z Q 2012 Numerical analysis of concrete temperature and stress in scouring sluice pier of gate dam during construction *Concrete* **22** 10-16 *(in Chinese)*

[20] Liu X H, Ma G, Chang X L, Zhuo W 2012 The refined numerical simulation of pipe cooling in mass concrete based on heat-fluid coupling method *Engineering Mechanics* **29**(8) 159-64 *(in Chinese)*

[21] Ma C C, Li S Y, Zhao L J, Liu C Y 2012 Research on the calculation method of superficial heat preservation of massive concrete structure *Journal of Northwest AF University (Natural Science)* **40** 217-23 *(in Chinese)*

[22] Gong S G, Xie G L *2004 Commands and Parametric Programming in ANSYS Beijing, China Machine Press*

[23] Zhang G X 2004 Improved Fem algorithm for calculating thermal field in heterogeneous material *Journal of Hydraulic Power* **56** 71-6 *(in Chinese)*

[24] Zhu B F 1994 The simulation analysis of multilayer concrete structure and layer algorithm *Journal of Hydraulic Power* **46**(3) 21-30 *(in Chinese)*

[25] Zhu B F 2006 Current situation and prospect of temperature control and cracking prevention technology for concrete dam *Journal of Hydraulic Engineering* **37**(12) 1424-32

Author

Hongyang Zhang, born in March, 1981, China

Current position, grades: lecturer.
University studies: Bachelor degree in Water resources and hydropower engineering at Zhengzhou university. Master degree in Hydraulic Structure Engineering at Zhengzhou university. Doctor degree in Hydraulic Structure Engineering at Hohai university.
Scientific interest: CAD modelling, model test, and simulation analysis.

Application of support vector machine in driving ranges prediction of pure electric vehicle with dual – energy storage system based on particle swarm algorithm

Shuang Du[1, 2], Chuncheng Zuo[1*]

[1]*College of Mechanical Science and Engineering, Jilin University, Changchun City, Jilin Province, China*

[2]*College of Engineering Technology, Jilin Agricultural University, Changchun City, Jilin Province, China*

Abstract

Driving ranges was a key factor that may affect the popularization and development of pure electric vehicle (PEV) with dual-energy storage system (DESS). It relied on neural network for its prediction. However, the prediction effect was not satisfactory due to local minimization, slow convergence rate, overfitting phenomenon and so on. In order to be more accurate in prediction, this paper introduced the Support Vector Regress (SVR) to the vehicle with parameters optimized by particle swarm optimization (PSO). Compare to BP neural network algorithm, PSO-SVR algorithm is more accurate and practical.

Keywords: pure electric vehicle with dual-energy storage system, particle swarm optimization algorithm, support vector machine, driving ranges

1 Introduction

Electric vehicles (EV) are more environmental-friendly with less noise and save more energy compared with fuel vehicles. But their driving ranges are a big problem, with the battery of EV as the primary obstacle. Therefore, it is significant to study the relationship between SOC (state of charge) of batteries and driving ranges.

In recent years, many experts and scholars propose the prediction methods based on non-linear theory. Li Dinggen, Deng Jie, et al study the relationship curves between SOC of batteries and driving ranges according to the energy equivalent principle between EV energy consumption and output energy of batteries [1]. Xu Wenle uses RBF neural network to predict driving ranges according to EV speed, temperature, current and voltage. It is proved to be feasible after the comparison between predicted data and actual data [2]. Zhang Wanxing establishes a SOC model of lithium iron phosphate battery through BP neural network and simulates driving ranges of the model in cruise software [3].

At present, the prediction of EV driving ranges by neural network is widely accepted. Neural network can map the complex nonlinear relationship and has strong nonlinear fitting ability. However, its learning speed is slow and samples should be of high quality. If there are a large amount of samples, the system may be in the state of local minimization and the generalization ability will be far from satisfactory [4-7]. Vanpik proposes a new machine learning method: support vector machine (SVM) [8]. It can address the nonlinear problem, cope with small samples, high dimension and local minimal phenomenon

[9]. So it points a direction for EV driving ranges prediction. The paper uses SVM to predict PEV with DESS driving ranges. As the prediction accuracy of SVM is closely related to training parameters, particle swarm is introduced to optimize the parameters. The results show that the accuracy rate of prediction based on PSO SVM is higher, therefore, it can apply in the prediction of EV driving ranges.

2 The principle of PSO algorithm to optimize parameters of SVM

2.1 SVM REGRESSION MODEL

The method of SVM is based on VC dimension theory of statistical learning and minimal principle of structural risk, in which the main idea of classification is to find the optimal classification plane and improve the accuracy rate of classification [10,11]. The regression is based on classification. It estimates the function through sample learning [12].

Suppose regression training sample set is: $A=\{(x_1, y_1), \dots(x_l, y_l)\}$, where, $x_i \in X = R^n$, $y_i \in Y = R$, i=1, 2, …, l. x_i refers to input variables, y_i refers to expected values corresponding x_i and l refers to the number of training samples. Regression function is described as:

$$f(x) = <w\,x> + b, (1)$$

where $w \in R^n$, it refers to weight vector. $b \in R$, it refers to s bias threshold. $<>$ refers to inner product operation. In order to ensure the generalization ability of SVM model,

---

* *Corresponding author* e-mail: dushuang234@sina.com

$\frac{1}{2}\|w\|^2$ should be minimized. Regression diagram of two-dimensional plane is shown in Figure 1, where dots refer to the training samples and the solid line refers to the ultra plane $f(x)$. The dotted lines H_1 and H_2 refer to the planes which are parallel to $f(x)$ and the distance between the dotted lines and $f(x)$ is ε. Regression aims at finding optimal solution of Equation (2).
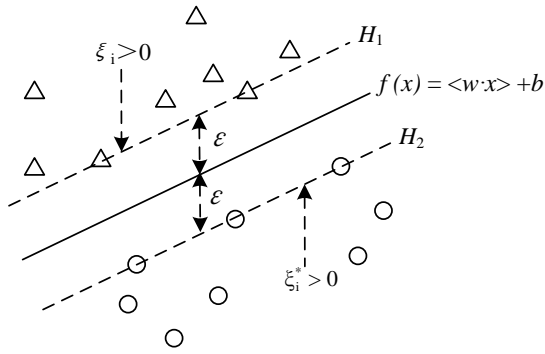

FIGURE 1 Diagram of SVR

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*), \tag{2}$$

s.t. $f(x_i) - <w\,x_i> - b \le \varepsilon + \xi_i,$

$<w\,x_i> + b - f(x_i) \le \varepsilon + \xi_i^*,$

$\xi_i, \xi_i^* \ge 0,$

$i=1,2, …, N,$

where C refers to penalty factor. ξ_i and ξ_i^* refer to slack variables. ε refers to insensitive loss coefficient. If there is the nonlinear relationship between the input variables and expected values, the kernel function will be introduced in SVM operation. Map input variables into high dimension space so the nonlinear problem in low dimensional space is shifted to a linear problem in high dimensional space [13,14]. In this paper, RBF is selected as the kernel function:

$$K(x_i, x_j) = \exp(\frac{-\|x_i - x_j\|^2}{2g^2}), \tag{3}$$

where g refers to the parameter of RBF kernel function. With Lagrangian function, the optimization of Equation (2) will be shifted to solving quadratic programming.

$$\max W(\alpha_i, \alpha_i^*) = -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) -$$

$$\varepsilon\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i,j=1}^{n}f(x_i)(\alpha_i - \alpha_i^*), \tag{4}$$

s.t. $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$, $0 \le \alpha_i \le C$ $i=1,2, … , N,$

$0 \le \alpha_i^* \le C,$

where α_i and α_i^* refer to Lagrangian multiplier.

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x). \tag{5}$$

So the Equation (1) can be expressed as:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x) + b. \tag{6}$$

2.3 THE BASIC PRINCIPLE OF PSO

PSO [15-17] is a new algorithm of swarm intelligence which is proposed by Kennedy and Eberhart. Its basic idea originates from the research on artificial life and birds preying behavior. The difference between PSO and the genetic algorithm is that PSO has not selection, crossover and mutation. It finds the optimal solution through searching the optimal solution of particles in the solution space. Suppose there are D particles in the swarm, the position vector of the i-th particle is $X_i = \{X_{i1}, X_{i2}, … , X_{id}\}^T$. The velocity vector of the i-th particle is $V_i = \{V_{i1}, V_{i2}, … , V_{id}\}^T$. The position vector of each particle corresponds to a fitness function value. In the iterative process, the particle updates own speed and position through tracking the optimal position of individual best position and global best position.

$$V_{id}^{k+1} = wV_{id}^k + c_1 r_1(P_{id}^k - X_{id}^k) + c_2 r_2(P_{gd}^k - X_{gd}^k), \tag{7}$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1}, \tag{8}$$

where P_{id}^k refers to the individual extremum of the k-th iteration. P_{gd}^k refers to the global extremum of the k-th iteration. c_1 and c_2 refer to acceleration constant. r_1 and r_2 are random numbers from zero to one. w refers to inertia weight.

2.3 PSO ALGORITHM FOR PARAMETERS OF SVM

In real situation, the choice of SVM parameters affects its performance to a great. The penalty factor C can define a Lagrange multiplier [18] and it also reflects the degree of punishment in the given feature space when the training model distributes erroneous sample data. The smaller the C value is, the smaller the punishment of sample data will be, but the bigger the training errors are. If the C value is greater, there will be more learning accuracy. But the generation ability of the model decreases. The appropriate C value can anti-interfere to a certain extent and enhance the stability of the model. g is the width coefficient of RBF kernel function which reflects the radial range of RBF core. If g value is smaller, the connection among SVM is loose and the generalization ability of model increases. If the g value is bigger, the number of support vectors will increase. But mode complexity also increases and is prone to produce under fitting [19]. Optimal methods for parameters of SVM are commonly used the grid search method [20,21], genetic algorithm and PSO. The paper uses PSO to optimize the parameters of SVM.

The steps of particle swarm algorithm to optimize the parameters of SVM are as follows:

1) Initialize the parameters C and g of PSO and determine the population size, then initialize the speed and position of the particle and set the number of iterations.

2) Regard individual optimal solution P_{ibest} of each particle as the current position and the current particle is substituted into the SVM model, and then calculate the fitness value of each particle. Regard the particle with maximum fitness value corresponding to individual optimal solution as the current global optimal solution P_{gbest}.

3) Update the position and speed of the particle according to the equation (7) and the equation (8).

4) Compare the fitness value of each particle with P_{ibest}. If the fitness value is better, P_{ibest} will be updated. Otherwise, the original value is retained.

5) Compare the fitness value of each particle with P_{gbest}. If the fitness value is better, P_{gbest} will be updated. Otherwise, the original value is retained.

6) Judge whether the system meets the exit conditions. If it does, the optimal fitness value will be returned. Otherwise, skip to step 2).

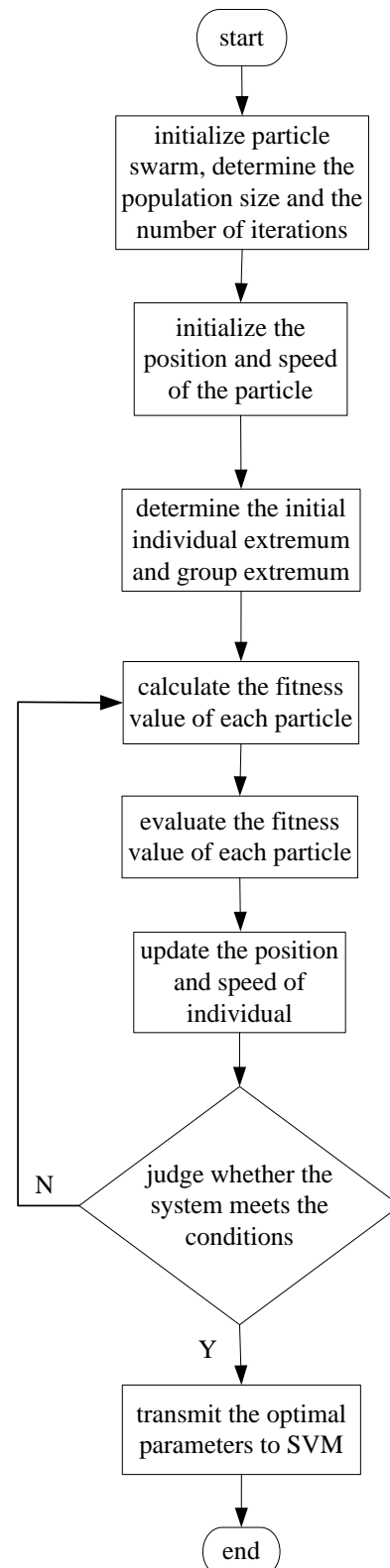Flow chart of the algorithm is shown in Figure 2.

FIGURE 2 Flow chart of PSO-SVR

3 Driving Ranges prediction of PEV with DESS

3.1 CALCULATING DRIVING RANGES

The driving ranges are one of the economic performances of EV. It refers to the distances from the completion of charge to the required standard. At present, EV driving ranges is calculated in the way that the output energy of batteries is equal to consumption energy of EV [22]. The driving force of EV:

$$F_t = mgf + \frac{C_D A v^2}{21.15} + \delta m j, \qquad (9)$$

where m refers to EV mass; f refers to the coefficient of rolling resistance; C_D refers to the coefficient of aero dynamic drag; A refers to frontal area; v refers to the driving speed of EV; δ refers to the coefficient of increased mass; j refers to the acceleration of EV. The expression of EV in the acceleration and deceleration is:

$$S = \frac{v^2}{2j}. \qquad (10)$$

The battery consumption energy is expressed as:

$$E_B = \frac{F_t S}{\eta \eta_M \eta_C \eta_B \times 3600}, \qquad (11)$$

where F_t refers to the driving force of EV; S refers to the driving ranges of EV; η refers to mechanical transmission efficiency; η_M refers to the efficiency of motor driving; η_C refers to the efficiency of motor controller; η_B refers to the efficiency of discharged batteries[23]. The parameters of PEV with DESS in the paper are shown in Table 1.

TABLE 1 The parameters of PEV with DESS

Parameter	Value	Parameter	Value
Vehicle Weight/kg	520	Mechanical transmission efficiency	0.92
Frontal Area/m²	1.05	Motor driving efficiency	0.93
Coefficient of Aero Dynamic Drag	0.35	Motor controller efficiency	0.96
Coefficient of Rolling Resistance	0.015	Discharged batteries efficiency	0.85

The driving conditions of EV include starting, acceleration, uniform speed, deceleration and stop. The driving distances of a complete driving condition is called driving section. The driving ranges are acquired by accumulated driving sections.

$$S = \sum_{i=1}^{k} S_i, \qquad (12)$$

where S_i refers to the driving distances of EV in each section; k refers to the number of driving condition [24]. The driving condition of EV in the paper is shown in Figure 3.
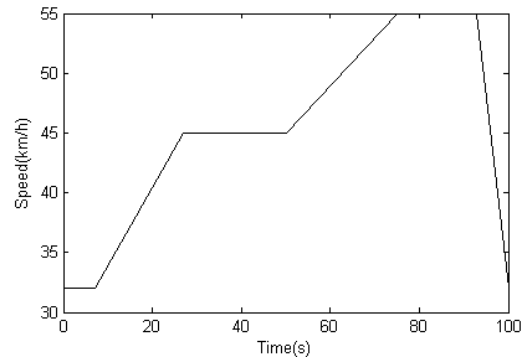

FIGURE 3 The cycle curve of EV

As we can see from Figure 3:

1) Electric vehicles drive at uniform speed from 0 to 7 seconds. Driving distances $S_1 = v_1 t = 62.23$m; driving force $F_{t1} = mgf + C_D A v /21.15 = 77.81$N; Energy consumption of batteries $E_{B1} = F_{t1} \times (S_1/1000)/\eta\eta_M\eta_C\eta_B \times 3600 = 0.0019$kw·h.

2) Electric vehicles accelerate from 7 to 27 seconds. Driving distances $S_2 = v_2/2j_1 = 432.78$m; driving force $F_{t2} = mgf + C_D A v / 21.15 + m \times j = 173.28$N; Energy consumption of batteries $E_{B2} = F_{t2} \times (S_2 / 1000) / \eta\eta_M\eta_C\eta_B \times 3600 = 0.0298$kw·h.

3) Electric vehicles drive at uniform speed from 27 to 50 seconds. The method is the same as 1). Driving distances $S_3 = 287.5$m; driving force $F_{t3} = 79.16$N; Energy consumption of batteries $E_{B3} = 0.0091$kw·h.

4) Electric vehicles accelerate from 50 to 75 seconds. The method is the same as 2). Driving distances $S_4 = 1051.57$m; driving force $F_{t4} = 134.57$N; energy consumption of batteries $E_{B4} = 0.0563$kw·h.

5) Electric vehicles drive at uniform speed from 75 to 93 seconds. Driving distances $S_5 = 275.02$m; Driving force $F_{t5} = 80.5$N; energy consumption of batteries $E_{B5} = 0.0088$kw·h.

6) Electric vehicles decelerate from 93 to 100 seconds. The driving distances $S_6 = v/2j_3 = 43.42$m.

The total energy consumption of batteries is 0.106 kw·h and driving ranges is 2152.52m. The driving time of EV is 9500 seconds. The driving ranges of EV in the driving condition is $S_{total} = 2152.52 \times 9500/100 = 204.49$km.

3.2 SAMPLE DATA AND PROCESSING

The energy storage devices of PEV with DESS include batteries and ultra-capacitor. There is a close connection between SOC changes of batteries and ultra-capacitor and driving ranges. Establish the model of PEV with DESS under Matlab/simulink. When the pure electric vehicles run in the normal state, the SOC change curves of batteries and ultra-capacitor are shown in Figure 4. The curve of driving ranges and time is shown in Figure 5. In general, the safe working ranges of batteries SOC are from 0.4 to 1 and the safe working ranges of ultra-capacitor SOC are from 0.2 to 1 [25]. As can be seen in Figure 4, 9500 seconds are needed when the SOC of batteries decreases to 0.4 and the SOC of ultra-capacitor decreases to 0.2.

Import the data of Figure 4 and Figure 5 into Matlab. The mutation data are removed according to the denoising experience then the smoothness of the signal is improved. According to the above, the prediction sample data are available. They are shown in Table 2.
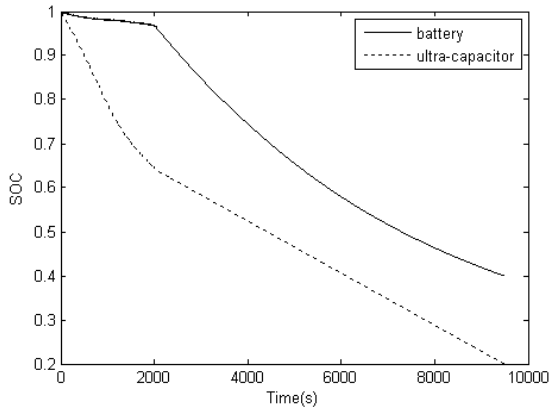

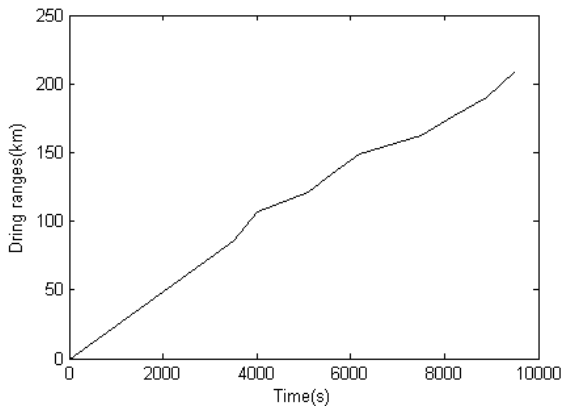FIGURE 4 The SOC-time curve of battery and ultra-capacitor

FIGURE 5 The curve of driving ranges and time

TABLE 2 The prediction sample data

NO.	Batteries SOC	Ultra-capacitor SOC	Driving ranges (km)
1	1	1	0
2	0.8212	0.5615	85.5521
3	0.7391	0.5117	106.9241
4	0.6656	0.4618	120.8139
5	0.6016	0.4115	134.7040
6	0.5464	0.3617	148.5938
7	0.4986	0.3113	162.4842
8	0.4572	0.2612	176.3737
9	0.4213	0.2116	190.2642
10	0.4000	0.2000	209.2100

From Table 2 it is clear that the driving range of EV which is measured through simulation experimental is 209.21km. It is close to 204.49km that is derived from theoretical calculation and the error is 2.3%.

3.3 THE PROGRAM OF PSO SVM ALGORITHM

PSO parameters of SVM are run in the Matlab environment. Establish the PSO model according to the Figure 2. The specific implementation programs are as follows:

```
load A.txt;        % import sample data
load B.txt;
c1 = 1.7;          % set the acceleration constant
c2 = 1.6;
maxgen=200;   % the number of iterations
sizepop=20;                        % population size
popcmax=10^(2);
% set the scope of optimal parameters C
popcmin=10^(–1);
popgmax=10^(3);
% set the scope of optimal parameters g
popgmin=10^(–2);
k = 0.6;
Vcmax = k*popcmax;        % set extreme speed of C
s=3;
%%% produce the initial particles and speed
for i=1:sizepop
pop(i,1) = (popcmax-popcmin)*rand+popcmin;
% generate initial population randomly
   pop(i,2) = (popgmax-popgmin)*rand+popgmin;
V(i,1)=Vcmax*rands(1);
% initialize speed
   V(i,2)=Vgmax*rands(1);
%%% calculate the initial fitness value
   cmd = ['-s',num2str(v),' -c ',num2str( pop(i,1) ),
'–g ',num2str( pop(i,2) )];
   fitness(i) = svmtrain(trainx, train, cmd);
fitness(i) = -fitness(i);
end
 %%% find extremum and extremum point
[global_fitness bestindex]=min(fitness);
% global extremum
local_fitness=fitness;
% initialize individual extremum
global_x=pop(bestindex,:);
% global extremum point
local_x=pop;
% initialize individual extremum point
```

Run the program of the above and the fitness curve is shown in Figure 6. As can be seen from Figure 6, the optimal fitness value is 0.91. The results show that the optimal parameters of SVM are $C = 2$ and $g = 0.25$ which are shown in Figure 7. The prediction results are evaluated by the mean square error and the square correlation coefficient. Mean square error:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (X_i - Y_i)^2 ,\qquad (13)$$

Square correlation coefficient:

$$r^2 = \frac{(N \sum_{i=1}^{N} X_i Y_i - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i)^2}{(N \sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2) \cdot (N \sum_{i=1}^{N} Y_i^2 - (\sum_{i=1}^{N} Y_i)^2)}, \quad (14)$$

where X_i refers to measured values. Y_i refers to prediction values. N refers to the number of prediction sample.

Establish the prediction model of SVM under the optimal parameters and predict the test data. The mean square error is calculated to be 0.005 according to Equation (13) and the square correlation coefficient is 0.98 according to Equation (14).

3.4 PREDICTION RESULTS AND ANALYSIS

Vcmin = -Vcmax ;
Vgmax = k*popgmax   % set extreme speed of g
Vgmin = −Vgmax ;
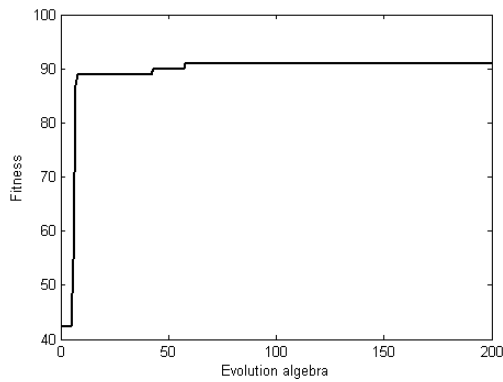
FIGURE 6 Fit curve of PSO

FIGURE 7 Diagram of C and g optimization results

BP neural network is a multilayer feed forward neural network. It is composed of input layer, hidden layer and output layer. There are two circulation signals between layer and layer, namely, the working signal and error signal. The working signal propagates forward after the input signal until the actual output signal at the output terminal. It is the function of the input signal and the weight. The error signal is the difference signal between actual output and expected output of neural network. It propagates back layer by layer through the output terminal [26,27].

Establish PSO-SVM algorithm prediction model according to the optimum parameters in section 2.3 and compare the prediction results between BP network model and PSO-SVM model.

From Figure 8, we can draw the conclusion that the change tends of SVM prediction, BP neural network prediction and the measured values is basically the same. But SVM prediction results are closer to the measured values. The prediction value of BP neural network has significant fluctuation. So its error is obvious. From Table 3, it is clear that the prediction results of SVM model are good. The maximum value of relative error of ten prediction samples is 1.91%. But the maximum value of relative error of BP neural network is 4.88%. So the prediction model of PSO parameters for SVM is reliable and reasonable and its accuracy rate is higher.
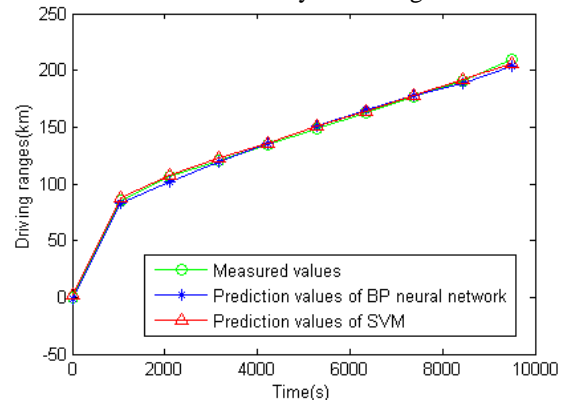
FIGURE 8 Results comparison diagram of between SVMand BP neural network driving ranges prediction

TABLE 3 Comparison of prediction results between SVR and BP neural network

NO.	Measured values	Prediction values of BP neural network model	Relative errors %	Prediction values of SVM model	Relative errors%
1	0.0102	2.6434	—	0.0082	—
2	85.5521	82.5505	-3.51	87.1834	1.91
3	106.9241	101.7097	-4.88	107.4154	0.46
4	120.8139	119.6689	-0.95	121.5251	0.59
5	134.7040	136.1486	1.07	135.7710	0.79
6	148.5938	151.1705	1.73	150.8951	1.55
7	162.4842	164.8946	1.48	163.9965	0.93
8	176.3737	177. 5508	0.67	177. 4479	0.61
9	190.2642	188.9525	-0.69	191.0705	0.42
10	209.2100	203.1256	-2.9	205.5227	-1.76

4 Conclusions

The paper proposes to predict the driving ranges of dual-energy PEV with SVM. It can overcome the local minimum phenomenon and overfitting phenomenon. At the same time, it can also address the uncertain hidden nodes of neural network.

The penalty coefficient and the width of radial basis function are optimized by particle swarm algorithm, then we establish SVM model. The comparison between SVM to BP neural network shows that the maximum relative error of SVM is less than 2%. But the maximum relative error of BP neural network is controlled within 5%. The results indicate that PSO parameters of SVM model have high fitting precision and are feasible for PEV with DESS to predict the driving ranges.

Acknowledgments

References

[1] Dinggen L, Jie D and Jun C 2008 A study of electric vehicle battery's state of charge estimation and its continued driving mileage *Industrial Instrumentation and Automation Device* (5) 91-4

[2] Wenle X 2012 Forecasting the driving range of pure electric vehicles based on RBF neural network *Transport of Portage College of Beijing Jiaotong University* Beijing.

[3] Wanxing Z 2012 Forecasting for power battery remaining capacity and driving range of electric vehicle *Vehicle engineering College of Hefei University of Technology*

[4] Taylor J W, Buizza R 2002 Neural network load forecasting with weather ensemble predictions *IEEE Transactions on Power Systems* **17**(3) 626-32

[5] Zhaoqi B, Gongxue Z 2010 *Pattern Recognition* Tsinghua University press: Beijing 73-89

[6] vanajakshi L, Rilett L R 2004 A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed *IEEE Intelligent Vehicles Symposium* USA: Texas 194-9

[7] Vapnik V, Levin E, Le C Y 1994 Measuring the VC-dimension of a learning machine *Neural computation* **6**(5), 851-76.

[8] Alhusaini A H, Prasanna V K, Raghavendra C S 2000 A framework for mapping with resource coallocation in heterogeneous computing systems *Proceedings of the 9th Heterogeneous Computing Workshop* 273-86

[9] Vapnik V N 1998 *Statistical learning theory* New York: *Wiley* 21

[10] Fen W, Tao M, Xu M 2012 Gold price forecasting model based on regression with support vector machine for particle swarm optimization *Journal of Lanzhou University of Technology* **39**(3) 65-9

[11] Cortes C, Vapnik V 1995 Support vector networks *Machine Learning* **20** 273-97

[12] Jian W, Yang Z and Rui H 2013 Fault detection and diagnosis of EMB sensor system based of SVR *Journal of Jilin University (Engineering and Technology Edition)* **43**(5) 1178-83

[13] Zhiliang M 2009 Applied research of mixture kernel support vector machine in modeling for fermentation process *Jiangnan University*

[14] Xuesong W, Yuhu C 2009 Theory, method and application of the machine learning *Science press: Beijing* 82-3

[15] Eberhart R, Kennedy J A new optimizer using particle swarm theory 1995 *Proceedings of the Sixth International Symposium on Micro Machine and Human Science* 39-43

[16] Kennedy J, Eberhart R 1995 particle swarm optimization 1995 *IEEE International Conference on Neural Networks* 1942-8.

[17] Poli R, Kennedy J, Blakwell T 2007 Particle swarm optimization *Swarm Intelligence* **1**(1) 33-57

[18] Donghui L, Jianpeng B, Ping F, Zhiqing L 2009 Study on the choice optimum parameters of support vector machine *Journal of Hebei University of Science and technology* **30**(1) 58-61

[19] Xu Z B, Dai M W, Meng D Y 2009 Fast and Efficient Strategies for Model Selection of Gaussian Support Vector Machine *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics* **39**(5) 1292-307

[20] Hongying L 2009 Research of kernels in support vector classifying Machine *Chongqing University*

[21] Guojun D, Limulati H, Maiti M, Hui Y 2009 Algorithms of optimizing SVM's kernel parameters with RBF kernel *Journal of Xinjiang University(Natural Science Edition)* **26**(3) 355-63

[22] Dkarimi M, Mokhtari H, IravaniM R 2000 Wavelet based online disturbance detection for power quality applications 2000 *IEEE Transactions on Power Delivery* **15**(14) 1212-20

[23] Yuliang T 2013 Analysis of performance and research of driving ranges of pure electric vehicle *Bus technology and research* (1) 9-11

[24] Guiming W 2010 Electric vehicle and its performance optimization *Machinery Industry Press: Beijing*

[25] Kaiwen Z 2012 Research on key technologies of the composite power based on the electric vehicles *Automatic College of Wuhan University of Technology*.

[26] Yun S 2010 Matlab implementation of BP neural network *Journal of Xiangnan University* **31**(5) 86-8

[27] Houlin L, Xianfang W, Yong W 2012 Power prediction for centrifugal pumps at shut off condition based on BP neural network 2012 *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)* **28**(11) 45-9

Authors

Shuang Du, born in August, 1977, Changchun, Jilin, China

Current position, grades: Ph.D. candidate at the College of Mechanical Science and Engineering at Jilin University and lecturer at the College of Engineering Technology at Jilin Agricultural University, China.
University studies: Bachelor and Master degrees in mechanical and electronic engineering from Changchun University of Technology.
Scientific interest: automatic control of electric vehicles.
Publications: 8 papers, including 3 teaching reform papers and 5 research papers.

Chuncheng Zuo, born in November, 1963, Fanshi, Shanxi, China

Current position, grades: Doctoral supervisor and professor at Jilin University.
University studies: Doctor degree in biological and agricultural engineering from Jilin University, Changchun, China.
Scientific interest: microfluidic batteries and new energy vehicles.
Publications: more than 100 papers.
Experience: Working at Jilin University since 1991.

Prediction of coal mine gas emission based on Markov chain improving IGA-BP model

Xiaoheng Yan[1, 2], Hua Fu[1], Weihua Chen[1*]

[1]*Department of Electrical and Control Engineering, Liaoning Technical University, Huludao Liaoning 125105, China*

[2]*Department of Safety Science and Engineering, Liaoning Technical University, Fuxin Liaoning 123000, China*

Received 2 June 2014, www.cmnt.lv

Abstract

There are a lot of factors that affect the gas emission, and among those there is a complicated and nonlinear relationship, so a BP neural network model based on immune genetic algorithm (IGA) was constructed to solve the problem of the traditional BP neural network such as, slow training speed, easy to be trapped into local optimums, and the premature convergence. In order to further improve accuracy of the prediction, the Markov chain was used to modify the residual series for the sample of bigger error. The correction result is more close to the measured value. The results showed that both the prediction accuracy and convergence speed of the IGA-BP model are better than the BP neural network model. The prediction after modified by Markov chain was further improved, the absolute average relative error of the prediction of the IGA-BP model is 2.40%.

Keywords: gas emission, immunity genetic algorithm, BP neural network, Markov chain, prediction

1 Introduction

Gas is one of the important factors threatening the safety production of mine. The situation of safety in production of coal mine is becoming more and more serious. The security incidents occurred in small and medium sized coal mines are particularly frequent. According to statistics, more than 80% of the accidents are related to gas emission. Therefore, accurate prediction of gas emission become more important [1].

The gas emission is affected by geological structure, coal seam, coal seam thickness and other factors, various factors mutual restraint, is a nonlinear dynamic system, time-varying, makes it difficult to accurately predict the [2] of gas emission. In recent years, some scholars use BP neural network prediction of mine gas emission, and achieved good results, but BP neural network has slow convergence speed, easily falling into local minimum problems [3].

In order to solve these problems, this paper presents a BP neural network optimization algorithm -- the improved immune genetic algorithm BP neural network algorithm (Immunity Genetic Algorithm-BP neural network, referred to as IGA-BP), and to establish prediction model of underground gas emission of working face of dynamic prediction.

Although the improved algorithm better prediction results, but some relative error between predicted value and measured value is too large, so this paper use Markoff chain modified prediction error residual value, the correction value is more close to the measured values of [4].

2 Immune Genetic Algorithms (IGA)

In recent years, people began to biological immune technology into the bionic algorithm in traditional artificial immune algorithm, resulting in [5]. The theory based on this algorithm with immune operator, immune vaccination and immune selection operator has two aspects, the former is based on the prior knowledge of individual genes in the improvement, improve the individual fitness; the latter is testing the vaccine individual after inoculation, the individual fitness increased retention, or eliminated. The improved immune algorithm is applied to BP neural network design, use of biological immune technology in the concentration mechanism and individual diversity retaining strategy of immune regulation, effectively solve the premature convergence problem of [6].

IGA algorithm mainly uses the following features [7]:

1) immune system) ability to generate antibody diversity, the differentiated cells, the immune system produces antibodies to different antigens in many. Use this function to maintain the individual diversity of the evolutionary process, improves the global search ability of the algorithm, to avoid falling into local optimal solution;

2) self regulatory mechanism, balancing mechanism of immune system by promoting and inhibiting antibody, quantity can be self regulating necessary antibody inhibition concentration, individual this correspondence in genetic algorithm and the promotion, can improve the local search ability of the algorithm;

3) immune memory function and antibody producing cells will retain some memory cells, for the same antigen into the future, the corresponding memory cells can be

---

[*] ***Corresponding author*** e-mail: fxlgd@163.com

rapidly stimulated and produced a large amount of antibody. Through this antigen memory recognition function in IGA algorithm, can speed up the search speed, improve the algorithm's search ability overall.

In general, the IGA algorithm is an improved genetic algorithm by biological immune mechanism is constructed, it will be practical to solve the problem of the objective function corresponds to the antigen, and the solution for the problem corresponding to the antibody [8]. The biological immune principle knowable, cell invasion of life body antigen immune system of division and differentiation, automatic produce antibodies to fight, this process was called immune response. In the process of immune responses, some antibodies as a memory cell to be preserved, when similar antigen invaded again, memory cells stimulated and rapidly produce large amounts of antibodies, the secondary response than the initial response more quickly, strong, reaction of the immune system and memory function. Combined with the antibody and antigen, destroys the antigen in a series of reaction. At the same time, mutual promotion and inhibition between antibody and antibody, used to keep the diversity of antibodies and immune balance, the balance is based on the concentration mechanism, namely antibody concentration is high, are inhibited; low concentration, is promoted by the reaction of the self-regulatory function, immune system.

Compared with the standard GA algorithm, based on the characteristics of the main IGA algorithm immune principles:

1) characterized by immune memory, this feature can improve the search speed, the ascension of the whole GA algorithm search function;

2) has the characteristics of antibody diversity, local search using this feature can improve GA algorithm;

3) has the ability of self-regulating, this function can improve the global GA algorithm search function, to avoid falling into local solution.

In short, the characteristics of IGA algorithm not only keeps the global GA algorithm parallel random search function, but also avoids the premature convergence problem in a certain extent, to ensure rapid convergence to the global optimal solution [9].

3 IGA-BP algorithm methodologies

Realization depends on BP neural network algorithm based on BP neural network IGA.

First, IGA receives an antigen (corresponding to a particular problem), followed by a set of randomly generated initial antibody (corresponding to the candidate solution); then calculates the fitness of each generation of the antibody, the antibody of crossover and mutation; groups through the concentration update strategy to generate the next generation of antibody group; until the termination condition is met, the algorithm ends. Which has the largest concentration of population is defined

fitness or close to the number of antibodies and antibody ratio of total maximum fitness is.

Specific Methods IGA BP neural network optimization algorithm is as follows.

3.1 ANTIBODY CODING

BP neural network learning is a continuous parameter optimization process repeatedly, if you use a binary encoding, will lead encoded string is too long, and in the time and you want to calculate the antibody decoded into a real number, which may affect the accuracy of network learning and algorithm running time. Therefore, the use of real-coded way, each antibody corresponds to one network structure, the number of hidden nodes to the right of re-mixed and network real-coded.

3.2 ANTIBODY FITNESS FUNCTION

Assuming an amount of the antibody population, each antibody corresponds to a network, as a combined network structure. Corresponding to the error of each antibody as an antigen of IGA Therefore, the fitness function of the antibody to the antigen [10]:

$$F(i) = \frac{1}{E_i + const} , \qquad (1)$$

$$E_i = \sum_p \sum_{out} (T_{p,out} - Y_{p,out})^2, (i = 1, 2, ..., n) , \qquad (2)$$

where *const* is greater than zero; $T_{p,out}$, $Y_{p,out}$ respectively the *p* training samples of the first two desired output and actual output; E_i antibodies corresponding to the *i* error.

3.3 GENETIC MANIPULATION

3.3.1 Cross

Use way cross at the intersection of two points, establish $X_1^l = [x_1^1, x_2^1, ..., x_n^1]$, $X_2^l = [x_1^2, x_2^2, ..., x_n^2]$ - the *l*-generation of two antibodies. In the *i* two arithmetic crossover point and *j* point, the next generation of antibodies is:

$$X_1^{l+1} = [x_1^1, ..., x_i', ..., x_j', x_{j+1}^1 ..., x_n^1]$$
$$X_2^{l+1} = [x_1^2, ..., x_i'', ..., x_j'', x_{j+1}^2, ..., x_n^2] , \qquad (3)$$

The x_k' and x_k'' ($i \le k \le j$) is produced by the following linear combination:

$$x_k' = \zeta x_k^1 + (1 - \zeta) x_k^2$$
$$x_k'' = \zeta x_k^2 + (1 - \zeta) x_k^1 , \qquad (4)$$

type in the ζ ratio, moreover $\zeta \in [0, 1]$.

3.2.2 The variation

Variation using the gauss mutation and the specific operation is the first antibody decoding into corresponding network structure. One by one according to the Equation (5) the ownership of the mutation network weight, mutated by hidden node components and weights to form a new antibody:

$$\Delta W = a\sqrt{F(i)}\mu(0,1) , \tag{5}$$

where: $a \in [0,1]$; $\mu(0,1)$ is Gaussian operators.

3.4 BASED ON THE CONCENTRATION OF UPDATES

Concentration group to update the overall goal is to restrain high levels of antibodies, while ensuring the selected probability high fitness individuals a greater concentration of antibody is C, adjust the chance for individual choice:

$$p(i) = \alpha C\left(1 - \frac{F(i)}{F_{\max}}\right) + \beta \frac{F(i)}{F_{\max}} , \tag{6}$$

where: $\alpha, \beta \in [0,1]$; $F_{\max}$ is the biggest fitness or close to the largest fitness antibodies.

From the Equation (6), the antibody concentration is higher, the fitness of antibody is likely to be selected is smaller. Antibody concentration, the smaller the fitness of antibody the greater the chance of being chosen.

3.5 BASIC STEP

Step 1 randomly generated antibodies in a certain weight distribution interval of initial population, and each antibody denotes a BP network.
Step 2 calculate the fitness of each antibody according to each antibody coding to draw corresponding BP network and its connection weights.
Step 3 will assess whether or not the performance of the BP network can converge to the desired accuracy range, if can output optimal BP networks, good execution Step 4 if not.
Step 4 genetic operation to antibody group, crossover and mutation.
Step 5 choice based on the concentration of antibody, produce a new generation of antibody group, jump to Step 2 continue to cycle.

4 The establishment of the absolute gas emission prediction model

This article selects several main influencing factors of absolute gas emission [11, 12] as the input layer of the model: Mining layer primitive gas content (p_1), buried depth of coal seam (p_2), thickness of coal seam (p_3), coal bed pitch (p_4), mining height (p_5), working face length (p_6), working face production rate (p_7), adjacent layers of primitive gas content (p_8), adjacent layer thickness (p_9), interlayer lithology (p_{10}), mining intensity (p_{11}). The combined gas emission as output layer of the model.

Model of the topological structure of BP neural network in the form of the 11-16-1 (as shown in Figure 1), a total of 28 the activation function of neurons in hidden layer neurons are s-shaped function:

$$f(x) = \frac{1}{1+e^{-x}} , \tag{7}$$

Output layer neuron activation function using linear function:

$$f(x) = x , \tag{8}$$

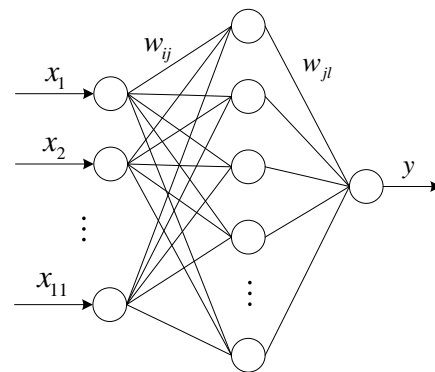

FIGURE 1 Topological structure of BF neural network

According to the characteristics of the immune genetic algorithm (GA), population size selected 50, type (1) = 0.001, (5) = 0.1, (4) = 0.5, the concentration of population in the total number of antibodies in the number and group is greater than 0.8, the ratio of evolution algebra for 200 generations.

5 Experimental studies

This paper selects kailuan mining group money mining camp in May 2007 to December 2008 working face of absolute gas emission quantity and influencing factors of statistical data (as shown in Table 1) as sample, 1~15 group of BP neural network model is used to IGA - training the BP model 16~18 group for testing, predictive results of the two models (see Figure 2) convergence contrast (as shown in Figure 3).

Through the experiment of Figures 2 and 3, the simulation curve can be seen that the BP neural network model of prediction in poor convergence speed, and the error is bigger; IGA - BP model is the improvement of BP neural network model, so the prediction curve can more accurately fitting the measured curve, precision and convergence speed are improved, but still part of the forecast error is bigger, so this article use Markov chain correction IGA - BP model prediction error of the salvage value, in order to improve the prediction accuracy.

493

TABLE l The statistical data of coalface gas emission influencing factors

Serial number	p_1 (m³/t)	p_2 (m)	p_3 (m)	p_4 (°)	p_5 (m)	P_6 (m)	P_7	P_8 (m³/t)	p_9 (m)	p_{10}	p_{11} (t/d)	T (m³/min)
1	1.92	408	2.0	10	2.0	155	0.960	2.02	1.50	5.03	1	3.34
2	2.15	411	2.0	8	2.0	140	0.950	2.10	1.21	4.87	1	2.97
3	2.14	420	1.8	11	1.8	175	0.950	2.64	1.62	4.75	1	3.56
4	2.58	432	2.3	10	2.3	145	0.950	2.40	1.48	4.91	2	3.62
5	2.40	456	2.2	15	2.2	160	0.940	2.55	1.75	4.63	2	4.17
6	3.22	516	2.8	13	2.8	180	0.930	2.21	1.72	4.78	2	4.60
7	2.80	527	2.5	17	2.5	180	0.940	2.81	1.81	4.51	1	4.92
8	3.35	531	2.9	9	2.9	165	0.930	1.88	1.42	1.82	2	4.78
9	3.61	550	2.9	12	2.9	155	0.920	2.12	1.60	4.83	2	5.23
10	3.68	563	3.0	11	3.0	175	0.940	3.11	1.46	4.53	2	5.56
11	4.21	590	5.9	8	5.9	170	0.795	3.40	1.50	4.77	3	7.24
12	4.03	604	6.2	9	6.2	180	0.812	3.15	1.80	4.70	3	7.80
13	4.80	634	6.5	9	6.1	165	0.785	3.02	1.74	4.62	3	7.68
14	4.80	634	6.5	12	6.5	175	0.773	2.98	1.92	4.55	3	8.51
15	4.67	640	6.3	11	6.3	175	0.802	2.56	1.75	4.60	3	7.95
16	2.43	450	2.7	12	2.2	160	0.950	2.00	1.70	4.84	1	4.06
17	3.16	544	2.7	11	2.7	165	0.930	2.30	1.80	4.90	2	4.92
18	4.62	629	6.4	13	6.4	170	0.803	3.35	1.61	4.63	3	8.04

6 Markov correction error salvage value

Markov chain is a kind of probability prediction method about the incident It is based on the status of the current events to predict its future changes each time a forecasting method of [13]:

1) Markov chain state Spaces is established.

2) To determine the state transition probability and the state transition matrix. In the process of development and change of events, the state E_i after k step move to the state transition probability of E_j to P_{ij} [14]:

$$P_{ij} = \frac{m_{ij}^{(k)}}{M_i}, \qquad (9)$$

where: A total number M_i state E_i; $m_{ij}^{(k)}$ state step E_i by k to E_j ; m is the division of state. Markov step m transition probability matrix in $P(m)$:

$$P(m) = \begin{bmatrix} P_{11}^{(m)} \cdots P_{1n}^{(m)} \\ \vdots \qquad \vdots \\ P_{n1}^{(m)} \cdots P_{nn}^{(m)} \end{bmatrix}. \qquad (10)$$
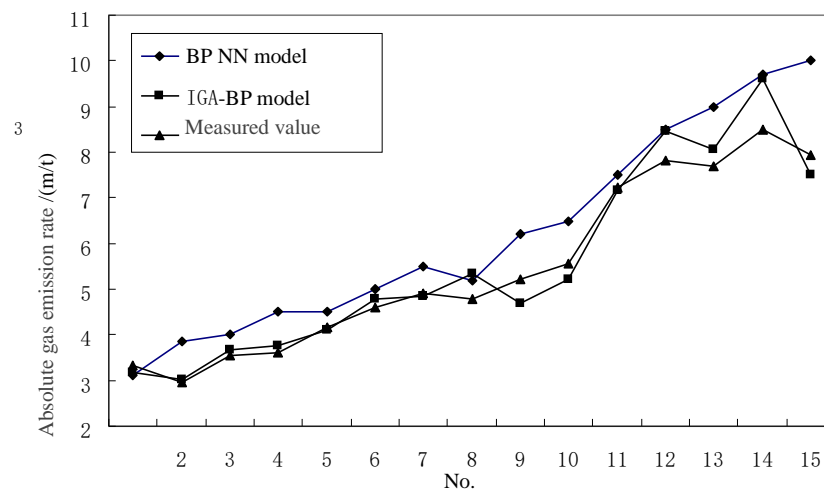

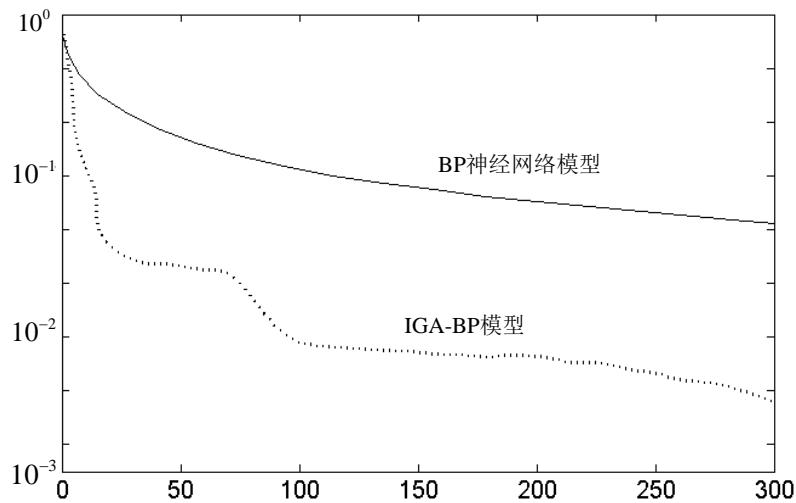FIGURE 2 The prediction results of BP neural network model and IGA-BP model

FIGURE 3 The convergence contrast of BP neural network model and IGA-BP model

3) IGA-BP model for Markov prediction results:

$$F = F_{IB}/(1-q),$$ (11)

where IGA - BP model prediction; to the original state of boundary value of the range according to the situation of training sample relative error of prediction (as shown in Table 2. Markov chain state interval is determined according to the actual situation can be divided into three intervals divided on the basis for $E_1[0 \quad \overline{x}-0.5s]$, $E_2[\overline{x}-0.5s \quad \overline{x}+0.5s]$, $E_3[\overline{x}+0.5s \quad 1]$ as sample mean square error.

Due to the hierarchical sequence must be positive, you must first make use of the equation:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}.$$ (12)

Normalized relative error sequence to [0, 1], between state interval.

According to the above formula of three state Markov state set respectively $E_1[0, 0.5458]$, $E_2[0.5458, 0.7021]$, $E_3[0.7021, 1]$ Using Equation (12) reduction divided interval, thus can be concluded that the relative error for sequence of three kinds of state. $E_1[-0.2892, -0.0988]$, $E_2[-0.0988, -0.0443]$, $E_3[-0.0443, 0.0596]$. According to the division of the state of the set of probability transfer matrix are obtained:

TABLE 2 The state of relative error distribution

Serial number	gas emission quantity(m³/t)		absolute error/relative error (%)	The normalized relative error (%)	Status
	measured value	*predictive value*			
1	3.34	3.19	4.54	83.05	E_3
2	2.97	3.01	1.32	82.88	E_3
3	3.56	3.66	2.83	82.84	E_3
4	3.62	3.75	3.54	82.82	E_3
5	4.17	4.10	1.66	82.97	E_3
6	4.60	4.79	4.09	82.80	E_3
7	4.92	4.86	1.29	82.96	E_3
8	4.78	5.35	11.85	82.58	E_1
9	5.23	4.70	5.96	83.09	E_3
10	5.56	5.23	4.91	83.06	E_3
11	7.24	7.17	0.95	82.95	E_3
12	7.80	8.46	8.42	82.68	E_2
13	7.68	8.05	4.77	82.78	E_2
14	8.51	9.61	12.87	82.55	E_1
15	7.95	7.50	5.60	83.08	E_3

TABLE 3 Prediction of IGA-BP model and Markov chain revised

Serial number	measured value /(m³/t)	IGA-B model		Markov	
		predictive value (m³/t)	*relative error (%)*	*modified value (m³/t)*	*relative error (%)*
16	4.06	4.26	-4.96	3.97	2.23
17	4.92	5.89	-19.63	4.96	-0.81
18	8.04	10.08	-25.36	8.47	-5.36

$$P(1) = \begin{bmatrix} 0.5714 & 0.1429 & 0.2857 \\ 0.6667 & 0.3333 & 0 \\ 0.0909 & 0.1818 & 0.7273 \end{bmatrix},$$

$$P(2) = \begin{bmatrix} 0.4477 & 0.1812 & 0.3711 \\ 0.6032 & 0.2063 & 0.1905 \\ 0.2393 & 0.2058 & 0.5549 \end{bmatrix},$$

$$P(3) = \begin{bmatrix} 0.4104 & 0.1918 & 0.3978 \\ 0.4996 & 0.1896 & 0.3109 \\ 0.3244 & 0.2037 & 0.4719 \end{bmatrix}.$$

So the Markov revised as shown in Table 3. As you can see from the result of modified Markov correction IGA - BP model can improve the prediction accuracy to make revised more close to the measured values With the actual and estimated values of the maximum relative error is 5.36%, the minimum relative error is 0.81%, the average relative

error was 2.40% to prove the author proposed method of residual error correction effect is obvious.

7 Conclusion

This paper presents a BP neural network based on IGA is applied in gas emission prediction IGA in Gaussian mutation genetic algorithm on the basis of introducing and updating strategy based on antibody concentration regulation mechanism, on the basis of effectively keep the diversity of antibodies, both retained the characteristics of global convergence of genetic algorithm, and can avoid the premature convergence problem in the very great degree, improves the convergence efficiency. By experimental verification show that IGA-BP model has a faster convergence speed and strong global convergence On this basis, the salvage value correction of Markov chain is applied to forecast error, the accuracy of IGA – prediction of BP model is further improved. Therefore, based on IGA-BP and Markov chain of coal mine gas emission prediction has certain application value.

References

[1] Wang T, Wang Y, Guo C 2012 Application of QGA-RBF for predicting the amount of mine gas emission *Chinese Journal of Sensors and Actuators* **25**(1) 120-3 *(in Chinese)*

[2] Pan Y, Deng Y, Zhang Q 2012 Gas emission prediction model based on QPSO-RBF *China Safety Science Journal* **22**(12) 29-34 *(in Chinese)*

[3] Huang J, Luo H, Wang H, et al 2009 Prediction of time sequence based on GA-BP neural net *Journal of University of Electronic Science and Technology of China* **38**(5) 687-92 *(in Chinese)*

[4] Fan Q, Li Wei, Wang Y et al 2011 An algorithm of shearer memory cutting based on grey-Markovian model *Journal of Central South University: Science and Technology* **42**(10) 3054-8 *(in Chinese)*

[5] Hong L, Mu Z 2006 Application of immune genetic algorithm in BP neural networks *Journal of University of Science and Technology Beijing* **28**(10) 997-1000 *(in Chinese)*

[6] Zhu Y, Zhang H, Su C 2009 Coal and gas outburst forecasting based on immune genetic algorithm *Journal of China University of Mining&Technology* **38**(1) 125-30 *(in Chinese)*

[7] Li X, Wang H, Xu C et al 2009 Calculation of line losses in distribution systems using artificial neural network aided by immune genetic algorithm *Power System Protection and Control* **37**(11) 36-9 *(in Chinese)*

[8] Duan Y, Ren W, Huo F 2005 A kind of new immune genetic algorithm and its application *Control and Decision* **20**(10) 1185-8 *(in Chinese)*

[9] Luo X, Wei W 2005 General discussion on convergence of immune genetic algorithm *Journal of Zhejiang University : Engineering Science* **39**(12) 2006-11 *(in Chinese)*

[10] Wang Q, Cheng J 2008 Forecast of coalmine gas concentration based on the immune neural network model *Journal of China Coal Society* **33**(6) 665-9 *(in Chinese)*

[11] Fu Hua, Jiang W, Shan X et al 2012 Study on coupling algorithm on coal mine gas emissionforecast model *Journal of China Coal Society* **37**(4) 654-8 *(in Chinese)*

[12] Zhu H, Chang W, Zhang B 2007 Different-source gas emission prediction model of working face based on BP artificial neural network and its application *Journal of China Coal Society* **32**(5) 504-8 *(in Chinese)*

[13] Xue P, Feng M, Xing X 2012 Water quality prediction model based on Markov chain improving gray neural network *Engineering Journal of Wuhan University* **45**(3) 319-24 *(in Chinese)*

[14] Tao Y, Xu J, Li S 2007 Predict gas emission quantity of mining coal face with improved gray Markov model *Journal of China Coal Society* **32**(4) 391-5 *(in Chinese)*

Authors

Xiaoheng Yan, born in October, 1984 ,Liaoning Anshan, China

Current position, grades: Lecturer at Liaoning engineering technology university institute of electrical and control engineering.
University studies: PhD student of the safety management of engineering at the university of Liaoning engineering technology university.
Scientific interest: security monitoring system, intelligent monitoring of coal mine gas.
Publications number or main: more than 10 papers, 2 educational materials.

Hua Fu, born in August, 1962, Liaoning Fuxin, China

Current position, grades: Professor, doctoral tutor at Liaoning engineering technology university institute of electrical and control engineering.
University studies: Postdoctoral graduate degrees.
Scientific interest: computer intelligent measurement, control technology and application, coal mine gas monitoring, research of modern sensing technology and system.

Weihua Chen, born in March, 1980, Heilongjiang Lanxi, China

Current position, grades: Lecturer at Liaoning engineering technology university institute of electrical and control engineering.
University studies: Doctoral student at Liaoning engineering technology university.
Scientific interest: Mechanical design and intelligent control, industrial automation production, control theory and control engineering research.
Publications number or main: 20 academic papers published, 1 educational material.

A computer aided kineto-static analysis method for spatial robot mechanism based on vector bond graph

Zhongshuang Wang[*], Yangyang Tao

School of Mechatronics Engineering, Qiqihar University, Qiqihar, Heilongjiang 161006, China

Abstract

In order to increase the reliability and efficiency of the kineto-static analysis of complex robot systems, the corresponding vector bond graph procedure is proposed. From the algebraic relations of input and output vectors in the basic fields, junction structure and Euler-junction structure of system vector bond graph model, the unified formulae of driving moment (or force) and constraint forces at joints are derived, which are easily derived on a computer in a complete form. For the algebraic difficulties brought by differential causality and nonlinear junction structure in system automatic modelling and kineto-static analysis, the effective bond graph augment method is proposed. Based on the kinematic constraint relations, the vector bond graph model of the spatial robot mechanism with five degrees of freedom can be made. As a result, the automatic modelling and kineto-static analysis of complex robot system on a computer is realized, its validity is illustrated.

Keywords: robot mechanism, kineto-static analysis, vector bond graph, causality, joint constrain

1 Introduction

The kineto-static analysis is very important for the control, static and dynamic strength check of robot system. For complex robot systems, e.g. the spatial robot systems containing different constraint joints, determining driving moment (or force) and the constraint forces at joints is a very tedious and error-prone task on account of the nonlinearities and couplings involved. The Newton-Euler technique and Lagrange technique are two of the well known methods used for the dynamic analysis of a robot system [1, 2]. These techniques however, are only suitable for a single energy domain systems, e.g. mechanical systems, and cannot be used to tackle systems that simultaneously include various physical domains in a unified manner.

The bond graph technique developed since the 1960's has potential applications in analysing such complex systems and has been used successfully in many areas [3, 4]. It is a pictorial representation of the dynamics of the system and clearly depicts the interaction between elements, it can also model multi-energy domains, for example, the actuator systems, which may be electrical, electro-magnetic, pneumatic, hydraulic or mechanical. But for spatial multibody systems such as spatial robot mechanism with different constraint joints, the kinematic and geometric constraints between bodies result in differential causality loop, and the nonlinear velocity relationship between the mass centre and an arbitrary point on a body leads to the nonlinear junction structure. The bond graph procedures mentioned above were found to be very difficult algebraically in automatic modelling and kineto-static analysis of system on a computer. To solve

this problem, the Lagrange multiplier approach and Karnopp-Margolis approach can be employed to model multibody systems based on scalar bond graph concept [5, 6].

For spatial multibody systems, the scalar bond graph technique is found to be complex and difficult. To address this problem, the vector bond graph techniques were proposed [7-9]. In vector bond graphs, single power bonds are replaced by multi-power bonds, this makes it posses more concise presentation manner and be more suitable for modelling spatial multibody systems. But some problems should be studied further, such as modelling spatial robot mechanism with different constraint joints by vector bond graphs, augmenting the vector bond model to avoid differential causality, developing the generic algorithm for automatic kineto-static analysis of spatial robot mechanism. To solve above problems, a more efficient and practical computer aided kineto-static analysis procedure for spatial robot mechanism based on vector bond graph is proposed here.

2 The vector bond graph model of spatial cylindrical joint

The diagram of spatial cylindrical joint is shown in Figure 1. This joint allows only a straight displacement and one direction rotation between its joined body B_α and B_β, fixing the remaining two translational and two rotational degrees of freedom. Therefore, only two generalized coordinates are free to change. Joint point P and Q are fixed on rigid body B_α and B_β respectively, vector h_α is used to describe the relative motion of the two rigid bodies,

---

[*] *Corresponding author* e-mail: wzhsh1962@163.com

$h_\alpha = r_\alpha^P - r_\beta^Q$. Where r_α^P and r_β^Q represent the position vector of joint point P and Q in global coordinates respectively, $r_\alpha^P = \begin{bmatrix} x_\alpha^P & y_\alpha^P & z_\alpha^P \end{bmatrix}^T$, $r_\beta^Q = \begin{bmatrix} x_\beta^Q & y_\beta^Q & z_\beta^Q \end{bmatrix}^T$. d_β^1 and d_β^2 are two unit vectors fixed on rigid body B_β, which are all orthogonal to slide axis, and orthogonal to each other. d_α is the unit vector fixed on rigid body B_α along slide axis, $d_\alpha^{'}$, $d_\beta^{'1}$ and $d_\beta^{'2}$ are the corresponding vectors in body frame. From the kinematic constraint condition of spatial cylindrical joint [1], we have
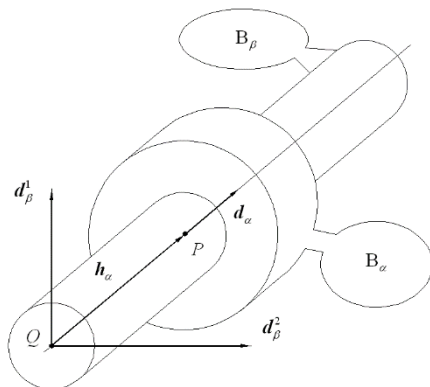
FIGURE 1 The diagram of spatial cylindrical joint

$$\underline{\Phi}^{(r_2)}(d_\alpha, d_\beta^1, d_\beta^2) = \begin{bmatrix} d_\beta^{1T} d_\alpha \\ d_\beta^{2T} d_\alpha \end{bmatrix} = \begin{bmatrix} d_\beta^{'1T} A^{\beta T} A^\alpha d_\alpha^{'} \\ d_\beta^{'2T} A^{\beta T} A^\alpha d_\alpha^{'} \end{bmatrix} = 0, \quad (1)$$

$$\underline{\Phi}^{(d_2)}(h, d_\beta^1, d_\beta^2) = \begin{bmatrix} d_\beta^{1T} h_\alpha \\ d_\beta^{2T} h_\alpha \end{bmatrix} = \begin{bmatrix} d_\beta^{'1T} A^{\beta T} \left(r_\beta^Q - r_\alpha^P \right) \\ d_\beta^{'2T} A^{\beta T} \left(r_\beta^Q - r_\alpha^P \right) \end{bmatrix} = 0, \quad (2)$$

where A^α and A^β are the direction cosine matrices of body B_α and body B_β respectively.

The corresponding velocity and angular velocity constraint equations can be written as:

$$\underline{\Phi}^{(r_2)}(d_\alpha, d_\beta^1, d_\beta^2) = \begin{bmatrix} d_\beta^{2T} \left(\omega_\alpha - \omega_\beta \right) \\ d_\beta^{1T} \left(\omega_\alpha - \omega_\beta \right) \end{bmatrix} =$$
$$\begin{bmatrix} d_\beta^{'2T} A^{\beta T} \left(\omega_\alpha - \omega_\beta \right) \\ d_\beta^{'1T} A^{\beta T} \left(\omega_\alpha - \omega_\beta \right) \end{bmatrix} = 0, \quad (3)$$

$$\underline{\Phi}^{(d_2)}(h, d_\beta^1, d_\beta^2) = \begin{bmatrix} d_\beta^{'2T} A^{\beta T} \left(\omega_\alpha - \omega_\beta \right) \\ d_\beta^{'1T} A^{\beta T} \left(\omega_\alpha - \omega_\beta \right) \end{bmatrix} = 0, \quad (4)$$

where ω_α and ω_β represent the angular velocity vectors of the rigid body B_α and B_β determined in global coordinates, ω_α^b and ω_β^b are the corresponding angular velocity vectors of the rigid body determined in body frame.

$$\tilde{d}_\beta^i = \begin{bmatrix} 0 & -d_{\beta z}^i & d_{\beta y}^i \\ d_{\beta z}^i & 0 & -d_{\beta x}^i \\ -d_{\beta y}^i & d_{\beta x}^i & 0 \end{bmatrix}, \ (i=1,2).$$

The velocity and angular velocity constraint equations shown as Equations (3) and (4) can be presented by vector bond model shown in Figure 2.
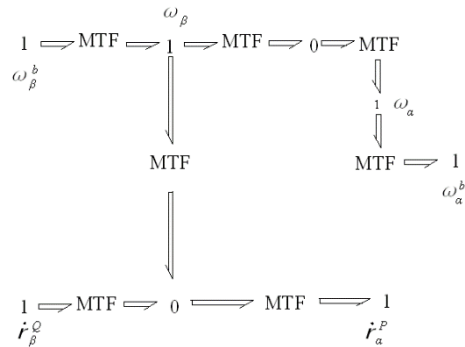

FIGURE 2 The vector bond graph model of spatial cylindrical joint

3 The unified formulae of driving moment and constraint forces for spatial robot systems

The basic fields and junction structure of system bond graph is shown in Figure 3 [3], where Euler-junction structure (EJS) [9, 10] is added. X_{i_1} represents energy vector variable of independent storage energy field corresponding to independent motion, X_{i_2} represents energy vector variable of independent storage energy field corresponding to dependent motion, Z_{i_1} and Z_{i_2} are the corresponding coenergy vector variables. D_{in} and D_{out} represent input and output vector variables in resistive field, U and V represent input and output vector variables of source field respectively, $U = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix}^T$, $V = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}^T$. Where U_1 is driving moment (or force) vector, U_2 is the constraint force vector of joint and U_3 is known source vector. E_{in} and E_{out} are the input and output vector variables in Euler-junction structure(EJS).
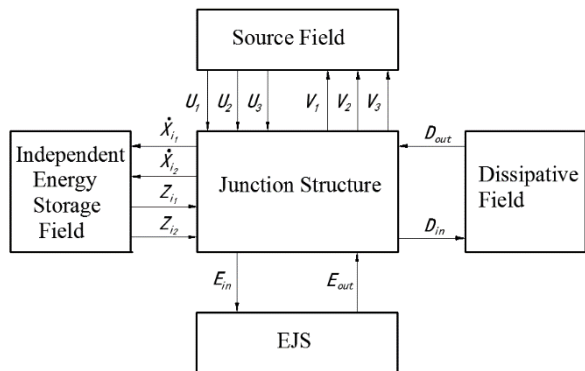

FIGURE 3 The basic field and junction structure of system

498

For independent energy storage field, we have:

$$Z_{i_1} = F_{i_1} X_{i_1} \,, \tag{5}$$

$$Z_{i_2} = F_{i_2} X_{i_2} \,, \tag{6}$$

where F_{i_1} and F_{i_2} are $m_1 \times m_1$ and $m_2 \times m_2$ matrices respectively.

For resistive field, we have:

$$D_{out} = R D_{in} \,, \tag{7}$$

where R is $L \times L$ matrix.

For Euler-junction structure (EJS), we have

$$E_{out} = R_E E_{in} \,, \tag{8}$$

where R_E is $L_E \times L_E$ matrix [9-11].

The corresponding junction structure equations can be written as:

$$\dot{X}_{i_1} = J_{i_1 i_1} Z_{i_1} + J_{i_1 i_2} Z_{i_2} + J_{i_1 L} D_{out} + \\ J_{i_1 u_1} U_1 + J_{i_1 u_2} U_2 + J_{i_1 u_3} U_3 + J_{i_1 E} E_{out} \,, \tag{9}$$

$$\dot{X}_{i_2} = J_{i_2 i_1} Z_{i_1} + J_{i_2 i_2} Z_{i_2} + J_{i_2 L} D_{out} + J_{i_2 u_1} U_1 + \\ J_{i_2 u_2} U_2 + J_{i_2 u_3} U_3 + J_{i_2 E} E_{out} \,, \tag{10}$$

$$D_{in} = J_{L i_1} Z_{i_1} + J_{L i_2} Z_{i_2} + J_{LL} D_{out} + J_{L u_1} U_1 + \\ J_{L u_2} U_2 + J_{L u_3} Z_3 + J_{LE} D_{out} \,, \tag{11}$$

$$E_{in} = J_{E i_1} Z_{i_1} + J_{E i_2} Z_{i_2} + J_{EL} D_{out} + J_{E u_1} U_1 + \\ J_{E u_2} U_2 + J_{E u_3} U_3 + J_{EE} E_{out} \,. \tag{12}$$

From the flow summation of 0-junctions corresponding to m_2 constraint force vectors in system vector bond graph model, we have:

$$0 = J_{C i_1} Z_{i_1} + J_{C i_2} Z_{i_2} + J_{CL} F_{out} + J_{C u_3} U_3 + J_{CE} E_{out} \,. \tag{13}$$

By the algebraic manipulation from Equations (5)-(13), the system driving moment and constraint force equations can be written as:

If $J_{CL} = 0$, $J_{CE} = 0$:

$$\left. \begin{aligned} U_1 &= S_{u_1 u_1}^{-1} (S_{u_1 i_1} X_{i_1} + S_{u_1 i_2} X_{i_2} + S_{u_1 u_3} U_3 + \\ &\quad T_{i_1 u_1}^{\mathrm{T}} \dot{X}_{i_1} + T_{i_1 u_1}^{\mathrm{T}} T_{i_1 u_2} H_4^{-1} J_{cu_3} \dot{U}_3) \quad (a) \\ U_2 &= (-H_4)^{-1} (H_1 X_{i_1} + H_2 X_{i_2} + H_3 U_1 + \\ &\quad H_5 U_3 + J_{cu_3} \dot{U}_3) \quad (b) \end{aligned} \right\}, \tag{14}$$

where:

$$A_1 = [I_2 - J_{EL} R (I_1 - J_{LL} R)^{-1} J_{LE} R_E - J_{EE} R_E]^{-1} \,,$$

$$A_2 = J_{E i_1} F_{i_1} + J_{EL} R (I_1 - J_{LL} R)^{-1} J_{L i_1} F_{i_1} \,,$$

$$A_3 = J_{E i_2} F_{i_2} + J_{EL} R (I_1 - J_{LL} R)^{-1} J_{L i_2} F_{i_2} \,,$$

$$A_4 = J_{L u_1} + J_{EL} R (I_1 - J_{LL} R)^{-1} J_{L u_1} \,,$$

$$A_5 = J_{E u_2} + J_{EL} R (I_1 - J_{LL} R)^{-1} J_{L u_2} \,,$$

$$A_6 = J_{E u_3} + J_{EL} R (I_1 - J_{LL} R)^{-1} J_{L u_3} \,,$$

$$B_1 = (I_1 - J_{LL} R)^{-1} (J_{L i_1} F_{i_1} + J_{LE} R_E A_1 A_2) \,,$$

$$B_2 = (I_1 - J_{LL} R)^{-1} (J_{L i_2} F_{i_2} + J_{LE} R_E A_1 A_3) \,,$$

$$B_3 = (I_1 - J_{LL} R)^{-1} (J_{L u_1} + J_{LE} R_E A_1 A_4) \,,$$

$$B_4 = (I_1 - J_{LL} R)^{-1} (J_{L u_2} + J_{LE} R_E A_1 A_5) \,,$$

$$B_5 = (I_1 - J_{LL} R)^{-1} (J_{L u_3} + J_{LE} R_E A_1 A_6) \,,$$

$$T_{i_1 i_1} = J_{i_1 i_1} F_{i_1} + J_{i_1 L} R B_1 + J_{i_1 E} R_E A_1 A_2 \,,$$

$$T_{i_1 i_2} = J_{i_1 i_2} F_{i_2} + J_{i_1 L} R B_2 + J_{i_1 E} R_E A_1 A_3 \,,$$

$$T_{i_1 u_1} = J_{i_1 L} R B_3 + J_{i_1 u_1} + J_{i_1 E} T_E A_1 A_4 \,,$$

$$T_{i_1 u_2} = J_{i_1 L} R B_4 + J_{i_1 u_2} + J_{i_1 E} T_E A_1 A_5 \,,$$

$$T_{i_1 u_3} = J_{i_1 L} R B_5 + J_{i_1 u_3} + J_{i_1 E} T_E A_1 A_6 \,,$$

$$H_1 = \dot{J}_{C i_1} F_{i_1} + J_{C i_1} F_{i_1} T_{i_1 i_1} + J_{C i_2} F_{i_2} T_{i_2 i_1} \,,$$

$$H_2 = \dot{J}_{C i_2} F_{i_2} + J_{C i_1} F_{i_1} T_{i_1 i_2} + J_{C i_2} F_{i_2} T_{i_2 i_2} \,,$$

$$H_3 = J_{C i_1} F_{i_1} T_{i_1 u_1} + J_{C i_2} F_{i_2} T_{i_2 u_1} \,,$$

$$H_4 = J_{C i_1} F_{i_1} T_{i_1 u_2} + J_{C i_2} F_{i_2} T_{i_2 u_2} \,,$$

$$H_5 = \dot{J}_{C u_3} + J_{C i_1} F_{i_1} T_{i_1 u_3} + J_{C i_2} F_{i_2} T_{i_2 u_3} \,,$$

$$S_{u_1 u_1} = T_{i_1 u_1}^{\mathrm{T}} [T_{i_1 u_1} + T_{i_1 u_2} (-H_4)^{-1} H_3] \,,$$

$$S_{u_1 i_1} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} H_4^{-1} H_1 - T_{i_1 i_1}) \,,$$

$$S_{u_1 i_2} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} H_4^{-1} H_2 - T_{i_1 i_2}) \,,$$

$$S_{u_1 u_3} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} H_4^{-1} H_5 - T_{i_1 u_3}) \,,$$

If $J_{CL} \neq 0$ or $J_{CE} \neq 0$:

$$\left. \begin{aligned} U_1 &= D_{u_1 u_1}^{-1} (D_{u_1 i_1} X_{i_1} + D_{u_1 i_2} X_{i_2} + D_{u_1 u_3} U_3 + \\ &\quad T_{i_1 u_1}^{\mathrm{T}} \dot{X}_{i_1}) \quad (a) \\ U_2 &= (-T_{C u_2})^{-1} (T_{C i_1} X_{i_1} + T_{C i_2} X_{i_2} + T_{C u_1} U_1 + \\ &\quad T_{C u_3} U_3) \quad (b) \end{aligned} \right\}, \tag{15}$$

where:

$$T_{C i_1} = J_{C i_1} F_{i_1} + J_{CL} R B_1 + J_{CE} R_E A_1 A_2 \,,$$

$$T_{C i_2} = J_{C i_2} F_{i_2} + J_{CL} R B_2 + J_{CE} T_E A_1 A_3 \,,$$

$$T_{C u_1} = J_{CL} R B_3 + J_{CE} T_E A_1 A_4 \,,$$

$$T_{C u_2} = J_{CL} R B_4 + J_{CE} R_E A_1 A_5 \,,$$

$$T_{C u_3} = J_{CL} R B_5 + J_{C u_3} + J_{CE} R_E A_1 A_6 \,,$$

$$D_{u_1 u_1} = T_{i_1 u_1}^{\mathrm{T}} [T_{i_1 u_1} + T_{i_1 u_2} (-T_{C u_2})^{-1} T_{C u_1}] \,,$$

$$D_{u_1 i_1} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} T_{C u_2}^{-1} T_{C i_1} - T_{i_1 i_1}) \,,$$

$$D_{u_1 i_2} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} T_{C u_2}^{-1} T_{C i_2} - T_{i_1 i_2}) \,,$$

$$D_{u_1 u_3} = T_{i_1 u_1}^{\mathrm{T}} (T_{i_1 u_2} T_{C u_2}^{-1} T_{C u_3} - T_{i_1 u_3}) \,.$$

Giving the system independent moving state variable vector X_{i_i} and its derivative $\dot{X}_{i_i}$, the corresponding system driving moment (or force) vector U_1 and constraint force vector U_2 can be determined from Equations (14) or (15) directly.

4 Example System

A robot mechanism with five degrees of freedom is shown in Figure 4, global coordinates $O_0X_0Y_0Z_0$ is located at the

point O_0, and the body frame $C_iX_iY_iZ_i$ ($i =1,2,3,4$) is located at the centre of mass. T_1, T_2 and T_3 are the driving moments, F_1 and F_2 are the driving forces along Z_1 axis and Y_2 axis respectively. The structure parameters of the robot mechanism are shown in Table 1, $c = 0.05$m, L=0.50m. The system input motion are as following, $Z_{C_1} = \cos(\pi t)$, $\theta_1 = \sin(\pi t)$, $Y_{C_2} = \cos(2\pi t)$, $\theta_2 = \sin(2\pi t)$, $\theta_3 = \sin(3\pi t)$.
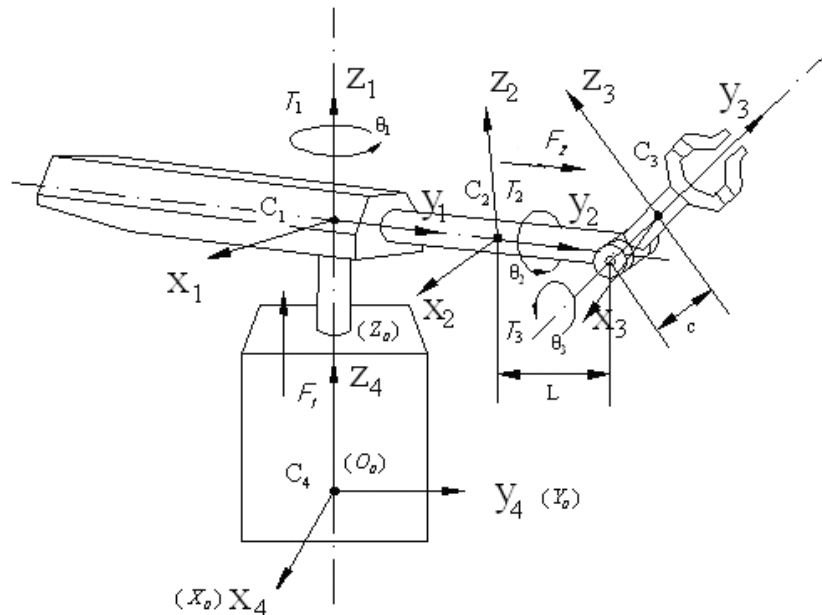
FIGURE 4 The robot system with five degrees of freedom

TABLE 1 The structural parameters of robot mechanism

Body i	Mass（kg）	Moment of inertia（Kg·m²）		
		X	Y	Z
1	250	90	10	90
2	150	13	0.75	13
3	100	4	1	4.3

The components for this example are four rigid bodies, which are joined by two cylindrical joints and one revolute joint, shown as Figure 5. For revolute joint, the constraint limits the relative translation of the two bodies B_α and B_β along three directions, and limits the relative rotation of the two bodies B_α and B_β along two directions, leaving only one rotation degree of freedom free. From the kinematic constraint condition, its vector bond graph can be obtained [11]. By the procedures mentioned above, the vector bond graph model of cylindrical joints can be made. By assembling the vector bond graph models of a single space moving rigid body [9, 11], the revolute joint,

and the cylindrical joints, the overall robot system vector bond graph model can be obtained and shown as Figure 6, where part I represents the cylindrical joint between body 1 and body 2, and part II is the revolute joint between body 2 and body 3.

Here, the constraint force vectors of joints can be considered as unknown source vectors, such as Se_3, Se_7, Se_{12}, Se_{14} in Figure 6 and added to the corresponding 0-junctions to eliminate differential causality. As a result, all differential causalities in this system vector bond graph can be eliminated, thus the procedure presented here can be used directly.
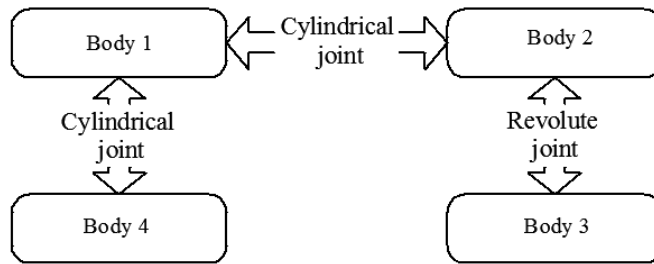
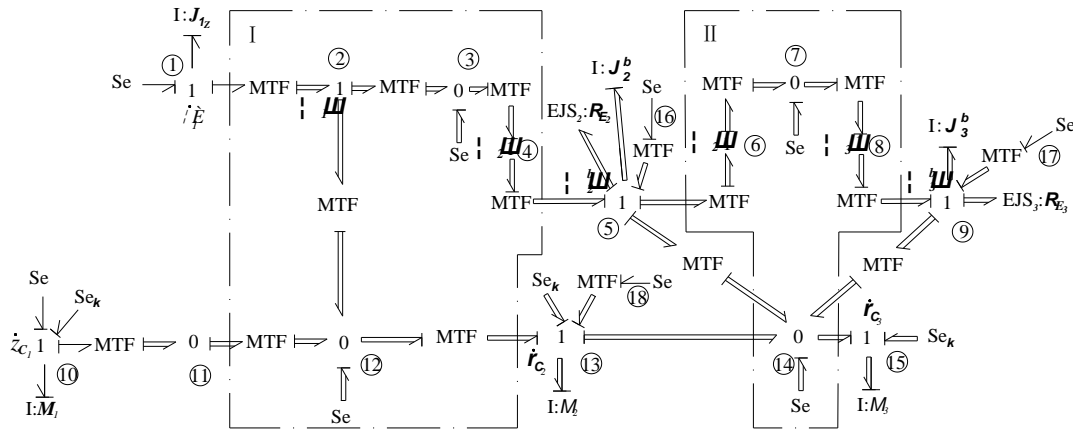FIGURE 5 The jointing structure diagram of robot system

FIGURE 6 The vector bond graph model of robot mechanism system

$J_{1_z} = [I_{Z_1}]$, $J_2^b = \mathrm{diag}(I_{X_2} \quad I_{Y_2} \quad I_{Z_2})$, $J_3^b = \mathrm{diag}(I_{X_3} \quad I_{Y_3} \quad I_{Z_3})$. $\dot{r}_{C_2}$, $\dot{r}_{C_3}$ are the mass centre velocity vector of body 2 and body 3 in global coordinates, ω_1^b , ω_2^b and ω_3^b are the angular velocity vector of body 1, body 2 and body 3 in body frame respectively, $\omega_1^b = \dot{\theta}_1$. ω_1 , ω_2 and ω_3 are the angular velocity vector of body 1, body 2 and body 3 in global coordinates. The mass of body i is m_{C_i} , $M_1 = m_{C_1}$, $M_2 = \mathrm{diag}(m_{C_2} \quad m_{C_2} \quad m_{C_2})$, $M_3 = \mathrm{diag}(m_{C_3} \quad m_{C_3} \quad m_{C_3})$.

Inputting the physical parameters of the robot mechanism, the coefficient matrices of Equations (5)-(13), known source vector U_3 , system independent moving state variable vector X_{i_1} , and its derivative $\dot{X}_{i_1}$ into the program associated with the procedure presented here based on MATLAB [12], the system driving moment (or force) and constraint force equations in the form of Equation (15) can be derived on a computer. The corresponding driving moment (or force) and constraint forces can be determined. Some of results are shown in Figures 7-10.
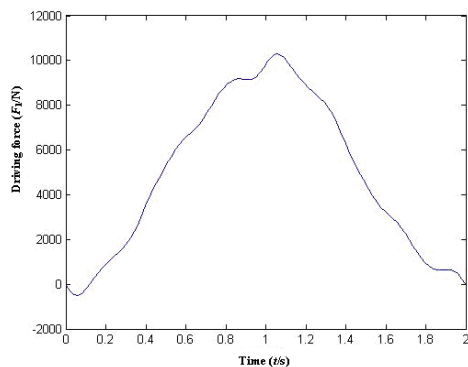

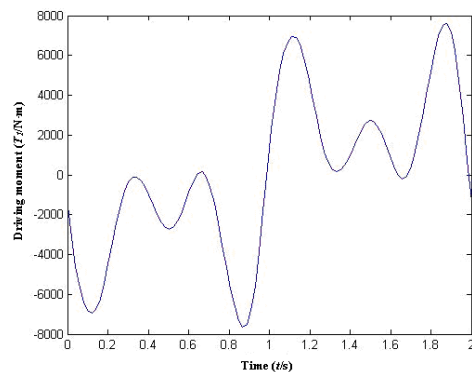FIGURE 7 The driving force of body 1
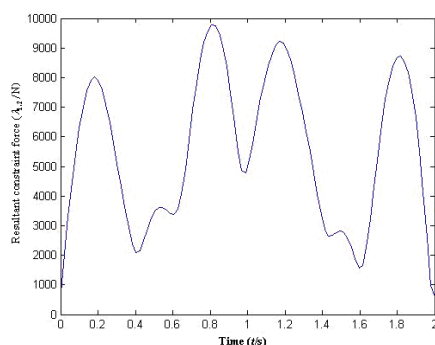

FIGURE 8 The driving moment of body 1

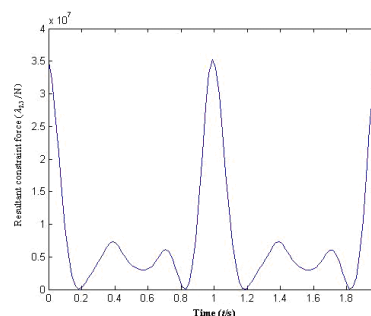FIGURE 9 Resultant constraint force between body 1 and body 2

FIGURE 10 Resultant constraint force between body 2 and body 3

5 Conclusions

The vector bond graph procedure presented here is very suitable for dealing with computer aided kineto-static analysis of complex robot systems with the coupling of multi-energy domains. Compared with traditional scalar bond graph method, this vector bond graph procedure is more suitable for complex spatial robot mechanism because of its more compact and concise representation manner. The differential causalities in the vector bond graph model of spatial robot mechanisms can be avoided by the bond graph augment method proposed here, thus the algebraic difficulties in system automatic modelling and kineto-static analysis can be overcome. In the case of considering EJS, the unified formulae of system driving moment and constraint force equations are derived, which

are easily derived on a computer in a complete form. These lead to a more efficient and practical automated procedure for kineto-static analysis of complex robot systems over a multi-energy domains in a unified manner. The validity of the procedure is illustrated by successful application to the kineto-static analysis of spatial robot systems with five degrees of freedom.

Acknowledgments

References

[1] Hong J Z 2003 Computational Dynamics of Multibody Systems *The Advanced Education Press* Beijing *(in Chinese)*
[2] Ma X F 1990 *Robot Mechanism Machine Press* Beijing *(in Chinese)*
[3] Karnopp D C, Margolis D L, Rosenberg R C 2006 System Dynamics: Modeling and Simulation of Mechatronic Systems (4th edition) *John Wiley* New York
[4] Kumar A, Pathak P M, Sukavanam N 2013 Trajectory Control of Two DOF Rigid-flexible Space Robot by A Virtual Space Vehicle 2013 *Robotics and Autonomous Systems* **61** 473-82
[5] Karnopp D C 1997 Understanding Multibody Dynamics Using Bond Graph Representations *Journal of Franklin Institute* **334B**(4) 631-42
[6] Margolis D L, Karnopp D C 1979 Bond Graph for Flexible Multibody Systems *Transaction of ASME Series G* **101**(1) 50-1
[7] Jinhee J, Changsoo H 1998 Proposion of a Modelling Method for Constrained Mechanical Systems Based on Vector Bond Graphs *Journal of the Franklin Institute* **335B**(3) 451-69

[8] Wilfird F, Serge S 1998 Bond Graph Representation of Multibody Systems with Kinematic Loops *Journal of the Franklin Institute* **335**(4) 643-50
[9] Behzadipour S, Khajepour A 2006 Causality in Vector Bond Graphs and Its Application to Modeling of Multi-body Dynamic Systems *Simulation Modelling Practice and Theory* **14** 279-95
[10] Breedveld P 2009 Stability of Rigid Rotation from a Bond Graph Pespective *Simulation Modelling Practice and Theory* **17** 92-106
[11] Wang Z S, Zou H, Tao Y Y 2013 Dynamic Modelling and Simulation of Electromechanical Systems Based on Vector Bond Graph 2013 *Advanced Materials Research* **706-708** 1294-9
[12] Jiang J F, Hu L J, Tang J 2004 Numerical Analysis and MATLAB Experiment *Science Press* Beijing *(in Chinese)*

Authors

Zhongshuang Wang, born in July, 1962, Shuang city, Heilongjiang, China

Current position, grades: Professor at Qiqihar University, master supervisor, academic leader of mechanical engineering.
University studies: Machinery Design and Manufacture in Qiqihar Light Industry Institute, Heilongjiang, China.
Scientific interest: multibody dynamics and control, robot dynamics.
Publications: 66 papers, 2 text books.
Experience: teacher's experience more than 16 years.

Yangyang Tao, born in November, 1986, Harbin, Heilongjiang, China

Current position, grades: Postgraduate of Mechanical Engineering at Qiqihar Univesity, Heilongjiang, China.
University studies: Food Science and Engineering in Zhejiang A&F University, Zhejiang, China.
Scientific interest: multibody dynamics and control, robot dynamics.
Publications: 4 papers.

The dynamic modelling and simulation for vehicle suspension systems based on vector bond graph

Zhongshuang Wang*, Zouqi Shu

School of Mechatronic Engineering, Qiqihar University, Qiqihar, Heilongjiang, 161006, China

Abstract

A dynamic modelling and simulation procedure for vehicle suspension system based on vector bond graph is described. According to kinematic constraint relations of components, the vector bond graph model for vehicle suspension system is built. In consideration of the coupling of energy variables and coenergy variables in independent energy storage field and dependent energy storage field. The unified formulae of system state space equations which are easily generated on a computer is derived. As a result, the automatic modelling and simulation of dynamics for vehicle suspension system can be realized based on MATLAB.

Keywords: vector bond graph, vehicle suspension system, modelling and simulation, mixed causality

1 Introduction

To increase the reliability and efficiency of the dynamic modelling and simulation of complex mechanical systems, different procedures have been proposed in previous work [1, 2]. But these procedures are only suitable for a single energy domain, such as strict mechanical one, and cannot be used to deal with the problems of computer aided dynamic analysis of mechanical systems containing the coupling of multi-energy domains. Bond graphs [3] have potential applications in analysing such complex systems because of their ability to describe the dynamics of interacting systems over a multi-energy domain in a unified manner and the unification of graph and mathematical descriptions. In many fields, bond graph techniques have been used successfully [4-6]. Compared with scalar bond graph [3], vector bond graph is more suitable for modelling complex systems such as vehicle suspension system because of its more concise representation manner [7-9]. In order to make the dynamic modeling and simulation of a system be carried out automatically on a computer, it is essential that a unified formula of system state space equations be derived. In what follows, the vector bond graph model for vehicle suspension system is built. In consideration of the coupling of energy variables and coenergy variables in independent energy storage field and dependent energy storage field, The unified formulae of system state space equations which are easily generated on a computer is derived. Finally, the automatic modelling and simulation of dynamics for vehicle suspension system can be realized based on MATLAB.

2 The vector bond graph model of vehicle suspension system

The vehicle suspension system is shown in Figure 1, it consists of vehicle body, passenger and seat system, front and rear suspensions, where v_{r_1} and v_{r_2} are the driving velocity of front and rear wheel from road. From kinematic analysis [1, 2], we have:

$$V_A = V_C + \begin{bmatrix} a\sin(q) \\ -a\cos(q) \end{bmatrix} \dot{q}, \tag{1}$$

$$V_B = V_C + \begin{bmatrix} -b\sin(q) \\ b\cos(q) \end{bmatrix} \dot{q}, \tag{2}$$

$$V_D = V_C + \begin{bmatrix} d\sin(q) \\ -d\cos(q) \end{bmatrix} \dot{q}, \tag{3}$$

where, q and $\dot{q}$ are the pitch angle and pitch angular velocity of the vehicle body, V_A, V_B and V_D are the velocities of point *A*, *B* and *D* on vehicle body, V_C is the mass centre velocity of vehicle body, *M* is the mass of vehicle body, and *J* is the rotational inertia of vehicle body. From Equations (1)-(3), the vector bond graph model of vehicle body can be obtained and shown in Figure 3, where the modulus matrices of MTF can be obtained from Equations (1)-(3) directly.

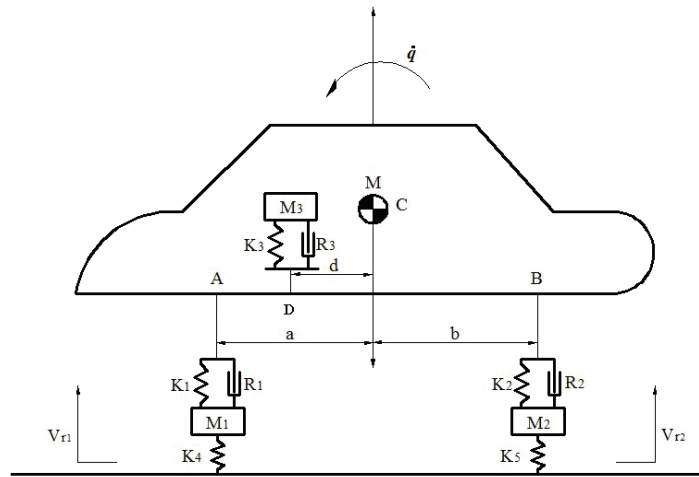* *Corresponding author* e-mail: wzhsh1962@163.com
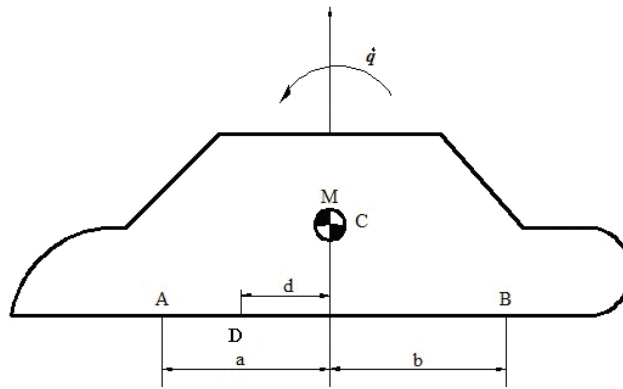
FIGURE 1 Vehicle suspension system
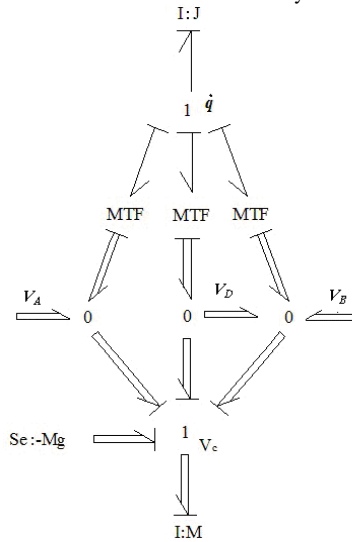

FIGURE 2 Vehicle body

FIGURE 3 The vector bond graph model of vehicle body

The passenger and seat system is shown in Figure 4, where M_3 is the mass of passenger. K_3 and R_3 are the spring coefficient and damper coefficient of seat respectively. Thus, the vector bond graph model of the passenger and seat system can be obtained and shown in Figure 5, where V_d is the velocity of passenger.
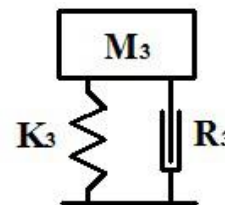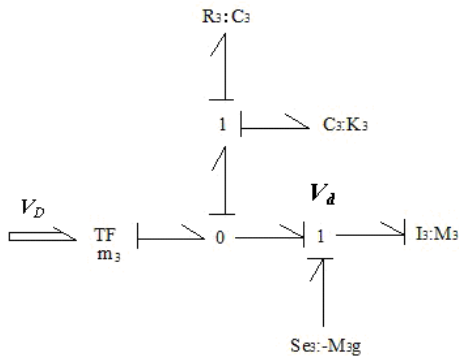

FIGURE 4 The passenger and seat system

504

FIGURE 5 The vector bond graph model of passenger and seat system

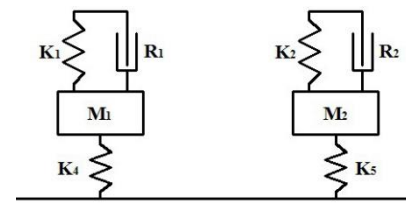The front and rear suspension systems are shown in Figure 6,

FIGURE 6 The front and rear suspension systems

where M_1 and M_2 are the mass of front wheel and rear wheel, K_1 and R_1 are the spring coefficient and damper coefficient of front suspension, K_2 and R_2 are the spring coefficient and damper coefficient of rear suspension, K_4 and K_5 are the spring coefficient of front tire and rear tire. Thus the vector bond graph model of the front and rear suspension systems can be made and shown in Figures 7a and 7b.
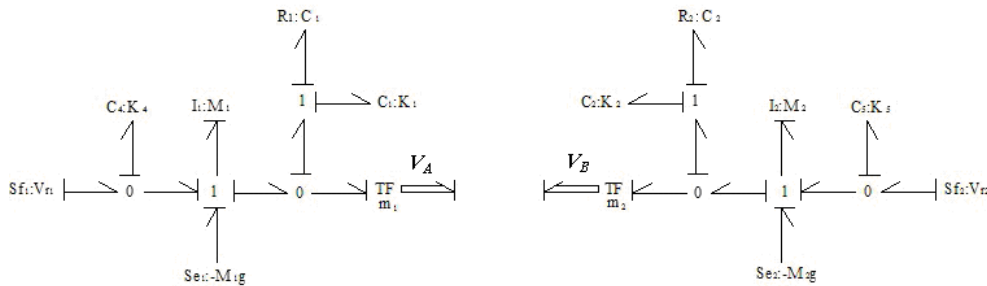

FIGURE 7 The vector bond graph model of front (a) and rear (b) suspension systems

By assembling the vector bond graph model of vehicle body, passenger and seat system, front and rear suspension systems, the overall vector bond graph model for vehicle suspension systems can be built and shown in Figure 8.

3 The unified formula of system state-space equations

The basic field and junction structure of system bond graph is shown in Figure 9, this is a basic field of system on which mixed causality is imposed [3]. In Figure 9, independent storage energy field consists of inertia element I and capacitance element C which posses integral causality, dependent storage energy field consists of inertia element I and capacitance element C which posses differential causality. Resistive field consists of resistance element R, and source field expresses the input to system from outer environment. Where X_i and X_d represent energy vector variables of independent and dependent storage energy field respectively, Z_i and Z_d are the corresponding coenergy vector variables respectively. D_{in}, D_{out} represent input and output vector variables in resistive field, U, V represent input and output vector variables of source field respectively.

For energy storage field, we have:

$$\begin{bmatrix} Z_i \\ Z_d \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} X_i \\ D_d \end{bmatrix}, \tag{4}$$

The coupling relations of vector variable X_i, X_d, Z_i and Z_d are expressed by Equation (4).

For resistive field, we have:

$$D_{out} = L D_{in}, \tag{5}$$
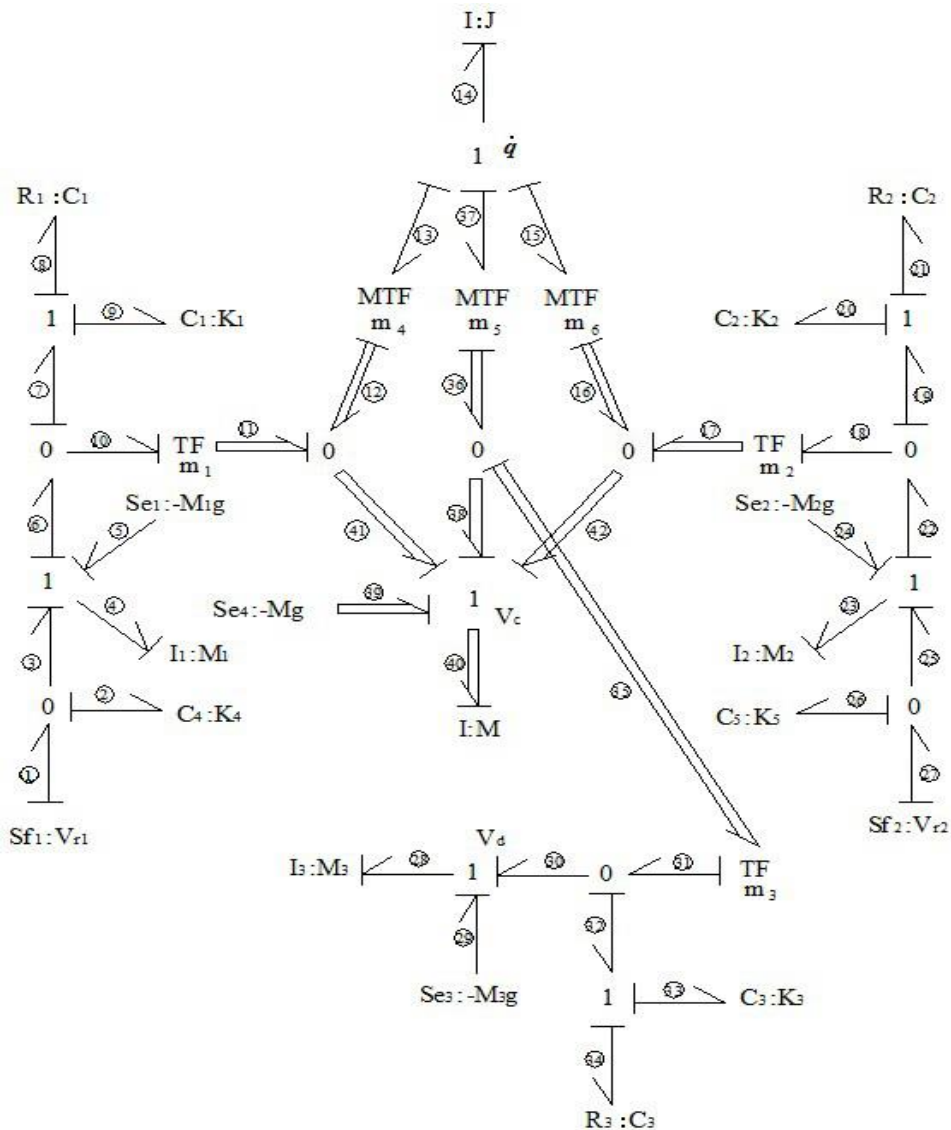
where R is a $L \times L$ matrix.

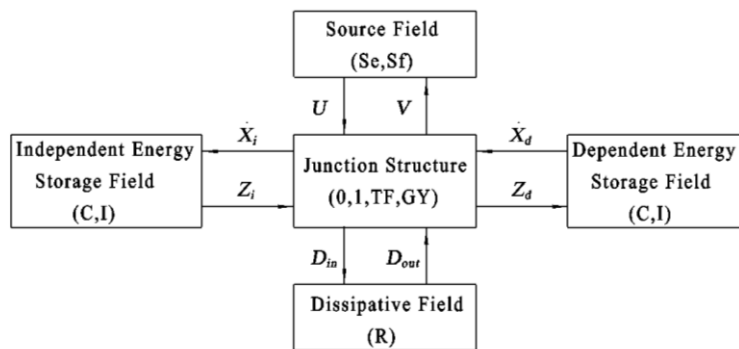FIGURE 8 The overall vector bond graph model of vehicle suspension system

FIGURE 9 The basic field and junction structure of system

The corresponding junction structure equations can be written as:

$$\dot{X}_i = J_i Z_i + J_{id}\dot{X}_{id} + J_{iL}D_{out} + J_{iu}U \ , \qquad (6)$$

$$Z_d = J_{di}Z_i + J_{du}U \ , \qquad (7)$$

$$D_{in} = J_{Li}Z_i + \dot{X}_d J_{Ld} + J_{LL}D_{out} + J_{Lu}U \ , \qquad (8)$$

By the algebraic manipulation from Equations (4)-(8), the system state space equations can be written as:

$$\dot{X}_i = AX_i + B_1 U + B_2\dot{U} \ , \qquad (9)$$

$$B_1 = T_1^{-1}T_3 \ , \ B_2 = T_1^{-1}T_4 \ , \ T_1 = I - J_{id}D - J_{iL}D_3 \ ,$$

$$T_2 = J_{ii}(F_{11} + F_{12}D) + J_{iL}D_2 \ ,$$

$T_3 = J_{ii}F_{12}D_1 + J_{iu} + J_{iL}D_4$,

$T_4 = J_{id}D_1 + J_{iL}D_5$,

$D = (I - F_{22}^{-1})^{-1}(F_{22}^{-1}J_{di}F_{11} - F_{22}^{-1}F_{21})$,

$D_1 = (I - F_{22}^{-1}J_{di}F_{12})^{-1}F_{22}^{-1}J_{du}$,

$D_2 = (I - LJ_{LL})^{-1}[LJ_{Li}(F_{11} + F_{12}D)]$,

$D_3 = (I - LJ_{LL})^{-1}LJ_{Ld}D$,

$D_4 = (I - LJ_{LL})^{-1}(LJ_{Li}F_{12}D_1 + LJ_{Lu})$,

$D_5 = (I - LJ_{LL})^{-1}LJ_{Ld}D_1$,

where I is a unit matrix.

For the system state space equations shown as Equations (9), many numerical solving algorithm that are available can be used. The corrected adaptive step size Runge-Kutta method based on MATLAB program [10] is explored here.

4 Example System

A vehicle suspension system is shown in Figure 1, the physical parameters are as followings, M=600 kg, M_1=35 kg, M_2=35 kg, M_3=80 kg, J=2500 kg·m², K_1=15700 N/m, K_2=15700 N/m, K_3=1000 N/m, K_4=120000 N/m, K_6=120000 N/m, R_1=1000 N·s/m, R_2=1000 N·s/m, R_3=500

N·s/m, a=1.4m, b=1.6m, d=0.8m. The driving velocities of front and rear wheels from road are as follows:

$$V_{r_1} = -0.057(t-3)\exp\left(\frac{-(t-3)2}{1.75}\right) \text{m/s},$$

$$V_{r_2} = -0.057(t-3-0.216)\exp\left(\frac{-(t-3-0.216)2}{1.75}\right) \text{m/s}.$$

From the vector bond graph model shown in Figure 8, the energy vector variables of independent and dependent storage energy field X_i and X_d , the corresponding coenergy vector variables Z_i and Z_d , the input and output vector variables in resistive field D_{in} and D_{out} , the input and output vector variables of source field U and V can be defined.

Inputting the initial values of state variable vector, the physical parameters of system, the coefficient matrices of Equations (4)-(8) into the program associated with the procedure presented here based on MATLAB [10], the system responses are obtained and shown in Figures 10-13.

For this example, the Newton-Euler method [1, 2] was used to determine the corresponding responses of the system, the results are in good agreement with that obtained by the procedure in this paper.
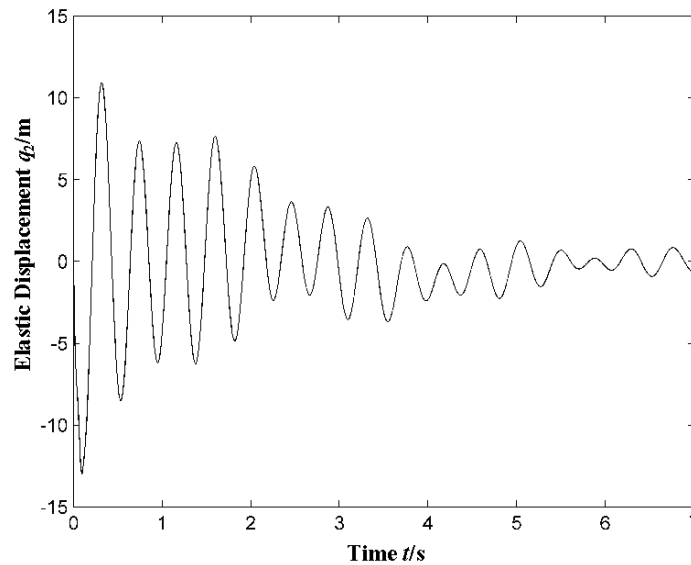

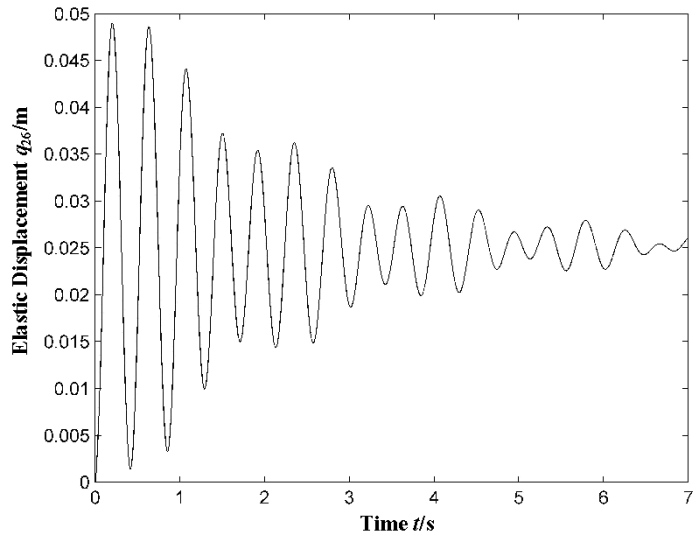FIGURE 10 The elastic displacement of front tire
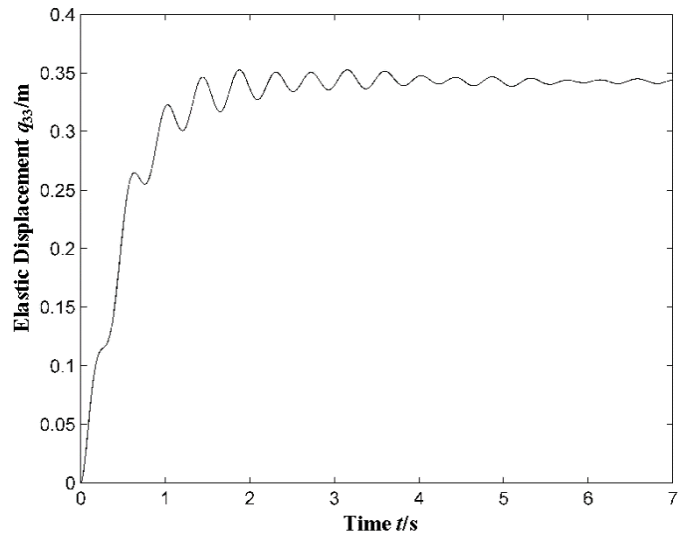
FIGURE 11 The elastic displacement of rear tire

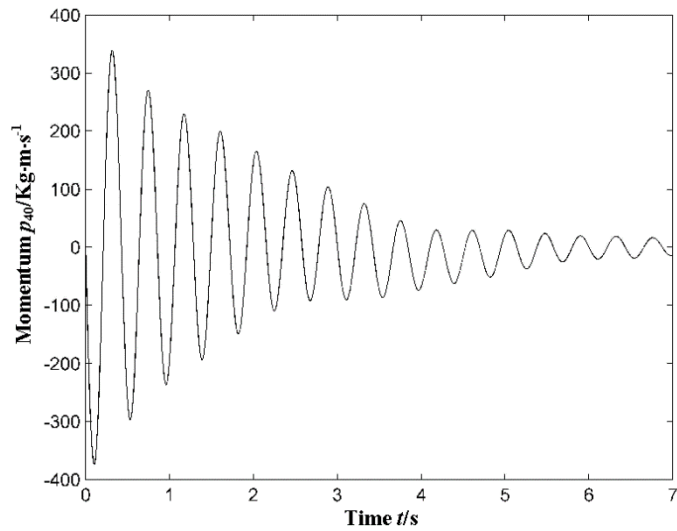FIGURE 12 The elastic displacement of passenger and seat system

FIGURE 13 The momentum of vehicle body in vertical direction

5 Conclusions

A vector bond graph procedure was presented for modelling and simulation of vehicle suspension system. Compared with standard scalar bond graph model, the procedure presented here is more compact and clear. In consideration of the coupling of energy variables and coenergy variables in independent energy storage field and dependent energy storage field, the unified formulae of system state space equations are derived. These lead to a more efficient and practical automated procedure for modelling and simulation of complex large-scale systems over a multi-energy domains in a unified manner. The validity of this procedure is illustrated by a practical example.

References

[1] Jiazhen H 2003 Computational Dynamics of Multibody Systems *China Advanced Education Press* Beijing *(in Chinese)*
[2] Xiangzhou Z, Guoyuan T 2010 Virtual Prototyping of Mechanical Systems *China Advanced Education Press* Beijing *(in Chinese)*
[3] Karnopp D C, Margolis D L, Rosenberg R C 2006 System Dynamics: Modeling and Simulation of Mechatronic Systems (4th edition) *John Wiley*: New York
[4] Chhabra R, Reza M 2011 Holistic System Modelling in Mechatronics *Mechatronics* **21** 166-75
[5] Bera T K, Bhattacharya K, Samantaray A K 2011 Evaluation of Antilock Braking System with an Integrated Model of Full Vehicle System Dynamics *Simulation Modelling Practice and Theory* **19** 2131-50
[6] Cipek M, Pavković D, Petrić J A 2013 Control-orientied Simulation Model of Power-split Hybrid Electric Vehicle *Application Energy* **101** 121-33
[7] Jinhee J, Changsoo H 1998 Proposion of a Modelling Method for Constrained Mechanical Systems Based on Vector Bond Graphs *Journal of the Franklin Institute* **335B**(3), 451-69
[8] Behzadipour S, Khajepour A 2006 Causality in Vector Bond Graphs and Its Application to Modeling of Multi-body Dynamic Systems *Simulation Modelling Practice and Theory* **14** 279-95
[9] Breedveld P 2009 Stability of Rigid Rotation from a Bond Graph Pespective *Simulation Modelling Practice and Theory* **17** 92-106
[10] Jiang J F, Hu L J, Tang J 2004 Numerical Analysis and MATLAB Experiment *Science Press* Beijing *(in Chinese)*

Authors

Zhongshuang Wang, born in July, 1962, Shuang city, Heilongjiang, China

Current position, grades: Professor at Qiqihar University, master supervisor, academic leader of mechanical engineering.
University studies: Machinery Design and Manufacture in Qiqihar Light Industry Institute, Heilongjiang, China.
Scientific interest: multibody dynamics and control, robot dynamics.
Publications: 66 papers, 2 text books.
Experience: teacher's experience more than 16 years.

Zuoqi Shu, born in February, 1987, Harbin, Heilongjiang, China

Current position, grades: Postgraduate of Mechanical Engineering in Qiqihar Univesity, Heilongjiang, China.
University studies: Vehicle engineering in Heilongjiang Institute of Technology, China.
Scientific interest: multibody dynamics and control, robot dynamics.
Publications: 1 paper

Optimal design of machine tool bed based on Ansys and orthogonal design

lei Wang[1, 2*]

[1]*Anhui Key Laboratory of Advanced Numerical Control and Servo Technology, Wuhu, 241000, P R China*

[2]*School of Mechanical and Automotive Engineering, Anhui Polytechnic University, Wuhu 241000, P R China*

Abstract

This paper attempts to use a simple and practical method based on the finite element theory and orthogonal design for the optimal structural parameters of machine tool bed. Dynamics finite element model of MCH63 machine tool bed has been established with the finite element analytic technique in order to improve the dynamic characteristics. The thickness of the ribs are looked as the design variables, the orthogonal design is used to optimize the natural frequency and the total mass. The simulation result indicates that the method is simple, effective and feasible.

Keywords: machine tool bed, optimal design, modal analysis, orthogonal design

1 Introduction

Machine tool bed is the major component of CNC machine tool, and its structural deformation affects machine tool's machining precision and dynamic characteristics greatly. With the ever-increasing demand in higher machine precision, the requirement for bed's stiffness is also increasing. At the same time, lightweight design is also being pursued for bed structure, which is important for a lower cost of material. However, it is difficult to obtain such a successful design due to the intricacy involved, including the diversity of layout pattern, the complexity of structural features as well as the variability of design parameters [1]. Recently, some researchers and scholars have done some work on machine tool bed [1-7]. However, there is much work to be done on CNC machine tool and its components in order to improve its stiffness, strength and dynamic characteristics.

As we know, the orthogonal design method is one of the experimental design methods. It samples some level combinations, which are orthogonal with each other, to represent the whole solution space. The selected combinations distribute uniformly in the solution space and more or less present some properties of their neighbours. The number of combinations to test is much smaller than the total number in solution space, and the combinations are representative and the results are convincing, so the experiments can be finished in a shorter time [8].This method can provide an effective way to find the near optimal solution combination of levels for the experiments, and it has been widely successfully applied in different areas [8, 9]. Therefore, it is more suitable for structural parameters of machine tool bed in order to obtain a better performance in this paper.

The remainder of this paper is organized as follows. Section 2 describes the model and modal analysis of the machine tool bed. The optimizing structural parameters based on orthogonal design are given in section 3. Concluding remarks are presented in section 4.

2 Modelling and modal analysis of the machine tool bed

The work space of MCH63 machine was designed to be $3768 \times 2080 \times 482 mm^3$. Bed material is HT300, and its density is $7350 kg/m^3$. The Poisson's ratio and elastic modulus are 0.27 and 1.3E11Pa respectively. A "well" structural layout for bed structure is formed in order to improve the stiffness, strength and dynamic characteristics. As a result, the simplified machine tool bed can be modelled, as shown in Figure 1. The finite element method (FEM) is popular and widely used in industrial area. Several studies have been carried out, modelling machine components with the FEM [10, 11]. Furthermore, the FEM is useful in the design process because many FEM software packages have useful interfaces to 3D CAD systems. Therefore, we also use FEM to process modal analysis for machine tool bed in this paper, and the finite element mesh is shown in Figure 2. By adding the boundary conditions to the finite element mesh model, the modes from the first order to the fourth order can be obtained, as shown in Figure 3-Figure 6. The natural frequencies from the first order to the fourth order are shown in Figure 7.

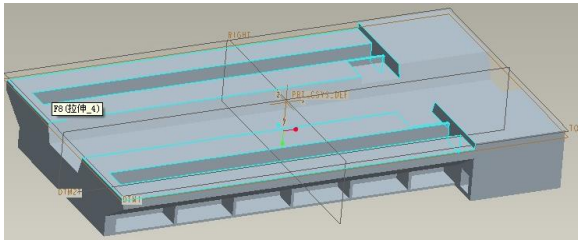[*]***Corresponding author*** e-mail: wangdalei2000@126.com
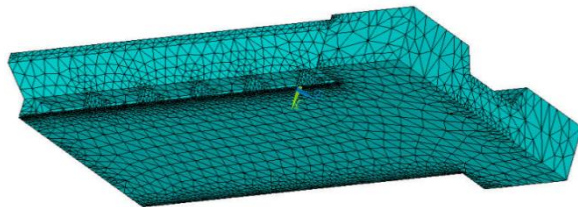
FIGURE 1 The simplified 3D model
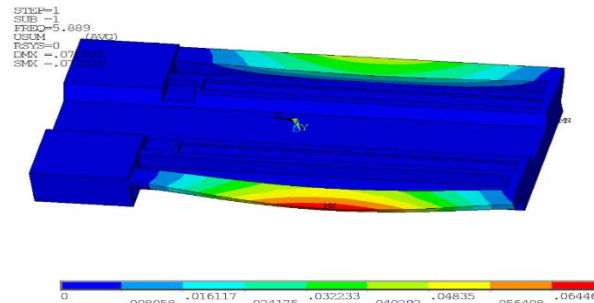

FIGURE 2 The finite element mesh

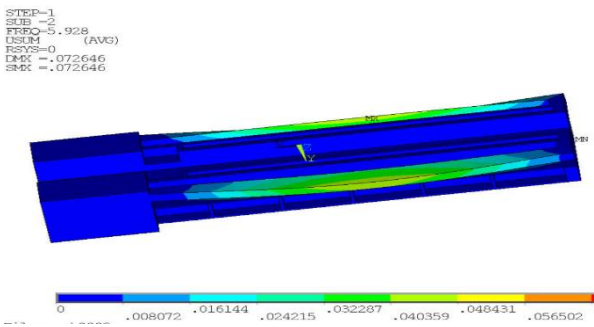FIGURE 3 The first order mode of the bed

FIGURE 4 The second order mode of the bed
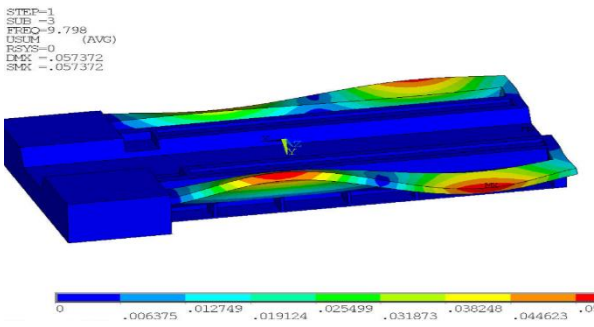

FIGURE 5 The third order mode of the bed
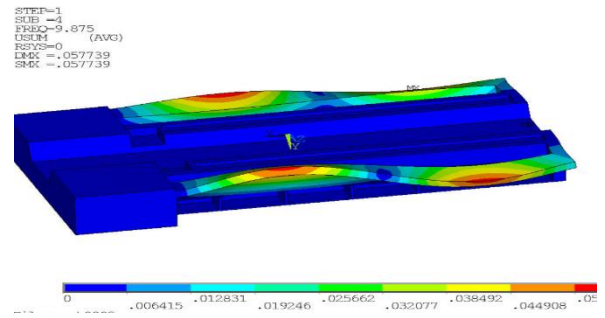
FIGURE 6 The fourth order mode of the bed

```
*****  INDEX OF DATA SETS ON RESULTS FILE  *****

SET     TIME/FREQ    LOAD STEP    SUBSTEP    CUMULATIVE
 1   5.8888            1            1           1
 File: prt0002

 2   5.9276            1            2           2
 File: prt0002

 3   9.7975            1            3           3
 File: prt0002

 4   9.8753            1            4           4
 File: prt0002

 5   14.310            1            5           5
 File: prt0002
```

FIGURE 7 Natural frequencies for each order mode

3 Optimal design based on orthogonal design

In this section, optimal selection of structural design parameters using orthogonal design for machine tool bed is given.

3.1 L25 (5⁶) ORTHOGONAL ARRAY EXPERIMENTS

The three five-level parameters, i.e. the thickness of the rib in the x direction (t_1), the thickness of the rib in the z direction (t_2) and the thickness of the rib on both sides (t_3) are considered. The optimal parameters and their selected levels are presented in Table 1. For three parameters at five levels each, the traditional full factorial design would require 5^6 experiments. However, in the current design (L25 (5^6) orthogonal array) the required experiments are only 25.

TABLE 1 Parameters and their respective levels in the present experimental design

Parameters	Level 1	Level 2	Level 3	Level 4	Level 5
t_1(mm)	22	24	25	26	28
t_2(mm)	26	28	29	30	32
t_3(mm)	22	24	25	26	28

The experimental lay out for genetic parameters using L25 (5^6) orthogonal array and the frequencies (f/Hz) and mass (M/Kg) are shown in Table 2. This design involves 25 separate experiments with specified levels and particular combination of parameters. However, the sequence in which the experiments were carried out was randomized to avoid any kind of

personal or subjective bias, which may be conscious or unconscious. This procedure ensures greater validity of test results [9].

3.2 EXPERIMENTAL RESULT AND ANALYSIS

As mentioned above, we use the natural frequency f and mass M as the evaluation indexes. As far as f is concerned, it for bigger-the-better characteristics is considered. The extreme difference analysis is carried out from Table 2, it can be obtained that the processing condition for attaining optimal f are t_1 at level 4, t_2 at level 2 and t_3 at level 2 (optimal scheme 1). In addition, we use the mass M as the evaluation index. As far as M is concerned, it for smaller-the-better characteristics is considered. The extreme difference analysis is carried out from Table 2, it can be obtained that the processing condition for attaining optimal f are t_1 at level 1, t_2 at level 1 and t_3 at level 1 (optimal scheme 2). The results are shown in Table 3. From the comparison results of these two kinds of optimal combination of parameters in Table 3, although the total mass M is a little bigger in optimal scheme 1 than that in optimal scheme 2, yet the natural frequency is improved in the optimal scheme 1.

Therefore, we select the optimal scheme 1 as the final result.

TABLE 3 Summarized results of frequency f and mass M

Scheme	Index	
	f/Hz	M/Kg
optimal scheme 1(t_1=26，t_2=28, t_3=24)	6.031	183.750
optimal scheme 2(t_1=22，t_2=26, t_3=22)	5.720	167.727

Finally, the 3D model for the machine bed tool by using the optimal combination parameters is shown in Figure 8.
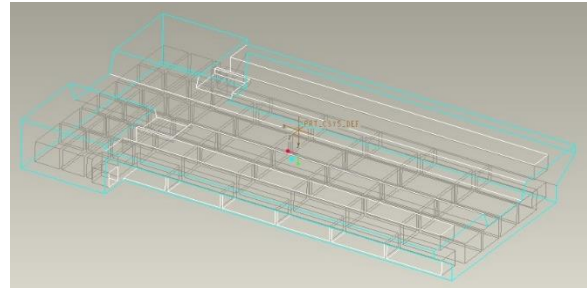

FIGURE 8 The optimal 3D model of the machine tool bed

TABLE 2 L25 (5^6) orthogonal array design and results

Experiments	Parameters			Index	
	t_1(mm)	t_2(mm)	t_3(mm)	f (Hz)	M(kg)
1	1(22)	1 (26)	1(22)	5.725	167.727
2	1(22)	2 (28)	2(24)	5.713	178.458
3	1(22)	3 (29)	3(25)	5.658	183.823
4	1(22)	4 (30)	4(26)	5.632	189.189
5	1(22)	5 (32)	5(28)	5.587	199.920
6	2(24)	1 (26)	2(24)	5.773	170.814
7	2(24)	2 (28)	3(25)	5.752	181.325
8	2(24)	3 (29)	4(26)	5.735	186.690
9	2(24)	4 (30)	5(28)	5.704	192.276
10	2(24)	5 (32)	1(22)	5.683	201.243
11	3(25)	1 (26)	3(25)	5.794	172.358
12	3(25)	2 (28)	4(26)	5.783	182.868
13	3(25)	3 (29)	5(28)	5.780	188.454
14	3(25)	4 (30)	1(22)	5.803	192.276
15	3(25)	5 (32)	2(24)	5.872	203.007
16	4(26)	1 (26)	4(26)	6.012	173.901
17	4(26)	2 (28)	5(28)	5.994	184.632
18	4(26)	3 (29)	1(22)	6.003	188.454
19	4(26)	4 (30)	2(24)	5.889	194.040
20	4(26)	5 (32)	3(25)	5.793	204.550
21	5(28)	1 (26)	5(28)	6.014	176.988
22	5(28)	2 (28)	1(22)	6.032	185.955
23	5(28)	3 (29)	2(24)	5.904	191.541
24	5(28)	4 (30)	3(25)	5.897	196.907
25	5(28)	5 (32)	4(26)	5.789	207.417

4 Conclusions

This paper attempts to develop a simple and practical method based on the finite element method and orthogonal design for the optimal structural design of machine tool bed in order to meet increasingly stiffness, strength and dynamic characteristics requirement. The proposed approach involves a two-phase procedure. Firstly, a simplified 3D model for machine tool bed is established.

With this model, the finite element method can be easily and economically employed to process modal analysis of the bed structure under actual operation conditions. Then, the thickness of the ribs are looked as the design variables, the orthogonal design is used to optimize the natural frequency and the total mass. Optimization results are finally elaborated to demonstrate the effectiveness of the proposed method.

References

[1] Li B, Hong J, Wang Z, Wu W, Chen Y 2012 *Procedia CIRP* **3** 626-31 *(in Chinese)*
[2] Liu S H, Ye W H, Chen W F, Yang Q 2010 *The Journal of Grey System* **22**(4) 341-9 *(in Chinese)*
[3] Wang J S, Tang X Q 2003 *International Journal of Machine Tools & Manufacture* **43**(7) 647-55
[4] Wang Z G, Cheng X, Nakamoto K, Kobayashi S, Yamazaki K 2010 *International Journal of Machine Tools & Manufacture* **50**(4) 57-65
[5] Daisuke K, Thomas L, Sascha W, Konrad W 2010 *Precision Engineering* **34**(3) 399-407
[6] Zeah M, Siedl D 2007 *Annals of the CIRP* **56**(1) 383-6 *(in Chinese)*
[7] Liu S H, Ye W H, Tang D B, Chen W F 2010 *Chinese High Technology Letters* **20**(9) 965-70
[8] Hu X M, Zhang J, Li Y 2008 *Journal of Computer Science and Technology* **23**(1) 2-18
[9] Puneet S, Amitabh V, Sidhu R K, Pandey O P 2005 *Journal of Materials Processing Technology* **168**(1) 147-51
[10] Zaeh M F, Oertli T H, Milberg J 2004 *Annals of the CIRP* **53**(1) 289-92 *(in Chinese)*
[11] Altintas Y, Cao Y 2005 *Annals of the CIRP* **54**(1) 379-82 *(in Chinese)*

Author

Wang lei, born on April 9, 1982, Bozhou, Anhui, China

Current position: an associate professor of Anhui Polytechnic University.
University studies: PH.D in 2010 from Nanjing University of Aeronautics and Astronautics, China.
Scientific interest: self-organization and self-adaptation manufacturing system, modelling and optimizing for manufacturing system, intelligent control theory and method for job shop scheduling.
Publications: more than 20 papers.

Evaluation of development patterns based on water resources carrying capacity calculation model in pastoral area

Yinhui Long, Zilong Liao*, Zhongxiao Guo, Yifan Song

Research Institute of Water Resources for Pastoral Area, Ministry of Water Resources, Hohhot, Inner Mongolia, China

Abstract

As the natural barrier, the health situation of grassland is relate to the change of social and economic development. According the natural environment of low rainfall and high evaporation, a comprehensive calculation model is established for evaluating the development patterns based on water resources carrying capacity. In present year, the social and grassland carrying capacity of BAILINGM is in the overload situation and there is a wide variance among different administrative region. By designing five different development patterns, we can find that carrying out the most strict water resources management policy can solve the overload problem of grassland carrying capacity effectively and the B-2 is the best development pattern. This paper show that the calculation results can provide reliable information and guide for macroscopic water resources management in large watershed and administrative region.

Keywords: development pattern, water resources, carrying capacity model, pastoral area

1 Introduction

China Pastoral area is mainly located in Inner Mongolia, Sinkiang, Qinghai and other province. It is easy to see worse nature conditions in these places, such as drought and less precipitation, higher frequent of sand and other natural disasters, more fragile ecological environment [1-3], so the quantity and quality of water resources can decide the suitable development of society and economy. Therefore, it is necessary to research the carrying capacity of water resources in pastoral area so that the protection of grassland ecology and environment can be strengthened.

2 Concepts and definition

2.1 CONCEPTS

The concept of water resources carrying capacity was introduced firstly by the teams of Shiyang Feng in the process of their investigating for water resources in Xinjiang Autonomous Region in 1992. The relative concept is given as "taking the change of society, economy and science technology as comprehensive index and selecting the population, the development and other factors as objective index [4]. Considering the principle of suitable development, Xu Xinyi pointed out that water resources optimization allocation is the key approach to improve water resources carrying capacity [5]. Then the expression of water resources carrying capacity was divided into direct evaluation and indirect evaluation by Ruan Benqing [6]. Jia Rong analysed the integration between the region and basin and the coordination between economy and ecology [7]. Wang Jianhua expanded the research into time and space scales [8]. Hui Yanghe considered the effects of both natural condition and human being activities [9].

2.2 DEFINITION

The concept of water resources carrying capacity should comply with the following requirement:

1) Space and time attribute. There are different grassland types, animal husbandry development level, the situation of water resources management, the grazing way with the change of water resources system in pastoral area.

2) Sustainable attribute. When the irrational development pattern surpass the recover ability of water resources, the water resources system will face the problems of quantity decreasing, poor quality and lower hydrology cycle stability.

3) Systematic attribute. The research of water resources carrying capacity is not only containing the subjective layer and objective layer, but also considering the coupling layer so that the relationship between the subjective layer and objective layer can be descried.

In some special period, taking the available water resources quantity as carrying subjective, considering the effects of history, policy, natural conditions, human activities and science technology, the water resources carrying capacity in pastoral area is described as the maximum carrying capacity of water resources system for supporting the balanced development of society, economy, ecology and environment.

---

* *Corresponding author* e-mail: zilong1115@qq.com

3 Basic theory

According the relationship between the carrying subjective and carrying objective, the calculation model based on supply-demand balance, which refer to the research fruit of Zuo Qiting [10], is established to calculate the water resources carrying capacity in pastoral area. The equations can be expressed as following:

$$\begin{cases} C_t = \varepsilon \cdot W_{\mathrm{p}} / E_p \\ C_{GDP} = W_e / E_{GDP} \\ C_g = (W_a - W_i - W_s) / E_g \\ W_r = W_a + W_e + W_t + W_p \end{cases},$$

where C_t is the capacity of water resources for supporting the population growth, ε is effective water using coefficient. E_p is total quota per person. C_{GDP} is the capacity of water resources for supporting the economy development. W_e is economic water demand quantity, E_{GDP} is GDP water consumption per ten thousand yuan. C_g is the capacity of water resources for supporting the grassland protection. W_r is total water resources quantity, W_a is first industry water consumption. W_p is living water consumption. W_e is second and third industry water consumption. W_t is ecology water consumption. W_i is farmland irrigation water quantity. W_s is animal husbandry water consumption. E_g is water using quota per farmland or artificial irrigation area.

4 Case study

4.1 GENERAL SITUATION

Damao County, which located in the middle-west of Inner Mongolia Province, is taken as studying area in this paper. Damao County is an area, which less precipitation and longer winter period, the average temperature per year is 4.34°C, the average precipitation per year is about 260.73mm, but the average evaporation is as high as 2480.57mm, and the evaporation value is 9.5 times than precipitation value. The grassland can be classify as four types: the fist type is type grassland which mainly including xerophytes herbs. The second type is desert grassland, which is less sparse distribution than the first and filled with xerophytes shrubs. The third type is grassland's desert that is filled with less bunch grass and is located in river basin. The last type is non-regional grassland among above three types. Damao County not only is a boundary county and animal husbandry base, at the same time, but also contains a lot of mineral resources. With the development of society and economy, the average disposable income in urban city has rose by 14.9 increasing rate per year, and the net income in village also has 12.5 increasing rate, and the industrial outcome has get to the level of the 10 billion RMB.

4.2 DEVELOPMENT PATTERNS

According to the national grassland protection construction development planning, national grassland ecological construction planning and the need of protecting typical grassland environment, natural grassland grazing level must be controlled within the appropriate scope of grazing capacity, the problem of excessive grazing must be prohibited so that the natural grassland can recover the original balance between ecology and environment. As we known, there is not a conflict between the development of ecology and society and the protecting of grassland environment, and the key solution is the development pattern of typical grassland. From the successful experience of the development patterns in the domestic and foreign countries. We can find that the suitable development road of grassland animal husbandry is decided by planning the scientific animal husbandry carrying capacity and choosing the appropriate grazing way, such as increasing higher production or better quality grass, improving the supply ability of artificial forage, developing the rotational, seasonal or regional grazing way, controlling the grazing frequency.

Under the above background, many problems will appear and we have to solve these difficulties, such as the protecting of grassland environment, population increasing, the change of grassland utilization way, the augment of pastoral economic aggregate, the improving of herdsman's net income, the irrigation scale of the artificial grassland.

Based on the practical situation of water resources conditions in typical grassland and the development need of economy and society in pastoral area, this paper mainly consider five different development patterns to forecast the future situation of pastoral area, and all the subsystem must be contain in the pattern in 2009 and 2019.

The present development pattern maintain the practical level in Damao County, and then the present water carrying capacity will be calculated for analysing the water consumer level and water using structure in different industries.

The planning year can be designed as following:

1) The subjective layer: assuming the initial water rights have been obtained, the available water resources use the present calculation results with considering the change of precipitation and evaporation.

2) The objective layer: in future water supplying conditions, the living water using and ecological water using should be satisfied first. In short, the living water using containing the third industry water using, the ecological water using including artificial and controlled water resources beyond the scale of river. Considering the vulnerable of ecology and environment conditions in pastoral area, the water supplying order of industry, animal husbandry and artificial grassland irrigation are superiority to farmland irrigation.

3) The coupling layer: water using efficiency and the net quantity of each unit are the primitive influence factors

515

in water resources coupling layer. Water using efficiency can be determined according to the development of industry standard. Based on the present standard, the net quantity of each unit will be improved with taking advance coefficient into account, other factors also be considered, such as the adjustment of agriculture farming structure, water saving irrigation upgrade.

TABLE 1 Typical Grassland Development Pattern in Planning Year

Pattern	Water Resources	Society	Economy	Ecology
A-1	$W=W_r$	$E_p(+1.0\%)$	$E_{GDP}(-2.0\%)$	$E_g(-1.0\%)$
A-2	$W=W_r$	$E_p(+1.0\%)$	$E_{GDP}(-5.0\%)$	$E_g(-2.0\%)$
B-1	$W=0.9 \cdot W_r$	$E_p(+0.8\%)$	$E_{GDP}(-2.0\%)$	$E_g(-1.0\%)$
B-2	$W=0.9 \cdot W_r$	$E_p(+0.8\%)$	$E_{GDP}(-5.0\%)$	$E_g(-1.0\%)$
C	$W=1.1 \cdot W_r$	$E_p(+1.0\%)$	$E_{GDP}(-2.0\%)$	$E_g(-1.0\%)$

W is the designed total available water resources quantity

4.3 RESULTS

4.3.1 Present year

The evaluation results of calculation model show that the water resources carrying capacity of supporting economic development in Bailingm Town and Mandul Town is overloading, because the distribution of the second and third industry in these two towns is intensive. Bailingm Town is the centre capital area of Damao County and its population density is high, so the population carrying capacity is overloading. However, in grassland ecology and environment aspect, the grassland carrying capacity of many towns is overloading expect the Bayinh Town and Mandul Town. Table 2 is the evaluation results of calculation model.

TABLE 2 Results of Water Resources Carrying Capacity Calculation in Present Year

Town	C_t	Present	Overload	C_{GDP}	Present	Overload	C_g	Present	Overload
	10 thousand people			*10 million RMB*			*km²*		
WUKE	4.4	3.1	-1.3	5.2	11.7	-40.6	882.2	984.4	102.2
SHIBAO	4.7	3.1	-1.6	56.0	8.2	-47.8	764.6	827.4	62.8
MINGAN	1.7	0.5	-1.2	20.6	8.0	-12.6	1591.9	1646.6	54.7
BAILINGM	2.7	3.6	0.8	32.6	58.3	25.7	2045.7	2846.4	800.7
BAYINH	0.8	0.4	-0.3	9.4	9.5	0.1	2289.2	2192.7	-96.4
MANDUL	0.3	0.2	-0.1	3.7	4.2	0.5	2330.7	1542.2	-788.5
DAERHAN	1.6	0.4	-0.1	19.1	5.4	-13.7	1799.9	3961.6	2161.6
XILAMR	1.6	0.5	-1.1	18.9	1.4	-17.5	987.8	1098.7	110.9
DAMAO County	14.4	12.0	-2.4	170.5	106.9	-63.7	12691.9	14500.0	1808.1

4.3.2 Planning year

Based on proper stocking rate and the natural rehabilitate by the recover ability of grassland itself, the income and the living level of herdsman should be considered the calculation model. When the grassland carrying capacity supported by water resources is in the overload situation, the surpass part can be supplied by artificial irrigation fodder grass.

The whole Damao County is taken as the average calculation unit of rainfall and economic development so that the administrative region is ignored. Forecasting the change of social and economic conditions, the carrying capacity of society, economy and grassland supported by water resources is be calculated and analysed in the planning year. The calculation results are shown in Table 3.

TABLE 3 Results of Water Resources Carrying Capacity Calculation in Planning Year

Patterns	C_t (10 thousand people)		C_{GDP} (10 thousand people)		C_g (km²)	
	Result	*Variant*	*Result*	*Variant*	*Result*	*Variant*
Present Year	12.0	0.0	106.9	0.0	14500.0	0.0
A-1	13.7	1.7	188.6	81.7	13346.0	-1154.0
A-2	13.8	1.8	220.3	113.4	14040.9	-459.1
B-1	13.1	1.1	209.3	102.4	13338.9	-1161.1
B-2	13.3	1.3	245.8	138.9	14787.5	287.5
C	15.7	3.7	242.4	135.5	14680.6	180.6

Compared with the present development patterns, the calculation results show that all carrying capacity of Damao County is in higher level. The *C* pattern is mostly improving the carrying capacity of society, economy and grassland due to the adding of water resources supplement. The *A*-1 and *A*-2 pattern, which is representing the current development situation is in higher social and economic carrying capacity but the grassland carrying capacity remain in the overload situation leaded by the irrational water demand structure and low ecological water available

quantity. By limiting the water using in first and second industry, the *Cg* of *B*-1 and *B*-2 pattern is up to 13338.9 and 14787.5 km² respectively. Although *C* pattern can solve the problem on excessive grazing activities, *B*-2 pattern is more suitable than *C* pattern in considering the management cost and technological feasibility. Therefore, putting on the strict water resources management policy is the most effective way to optimize the distribution of water resources supplement and comprehensive carrying capacity supported by water resources systems.

Long Yinhui, Liao Zilong, Guo Zhongxiao, Song Yifan

5 Conclusions

5.1 GENERAL SITUATION

Analysing the above calculation results, some conclusions can be summed as following:

1) The results of present year show that there is a wide variance among different administrative region, the water resources carrying capacity of supporting economic development in Bailingm Town and Mandul Town is overloading. And the population carrying capacity of Bailingm Town is overloading. In ecology and environment aspect, the grassland carrying capacity of many towns is overloading. Lower water resources carrying capacity is meaning the little water utilization effective, and the practice value of water resource, which is very limited, has not been revealed.

2) Damao County is located in the north of China, and its characteristic is low rainfall and high evaporation. Because of lacking the controlling the project of surface water, improving water resources system carrying capacity itself is the only way to remiss the overload situation. Considering the current development situation in pastoral area, five different development patterns are designed for

representing the social, economic and ecology conditions. As the natural barrier, the health situation of grassland is relate to the change of social and economic development. By calculating five different development patterns, the results show that *B*-2 pattern, which is designed for carrying out the most strict water resources management policy, can solve the overload problem of grassland carrying capacity compared with other development patterns.

3) From the research work, we can see that the calculation model of water resources carrying capacity is available and feasible at the aspect of management cost and technology, and the calculation results can provide reliable information and guide for macroscopic water resources management in large watershed and administrative region.

Acknowledgments

References

[1] Lu A, Ding Y, Pang H, Yuan L, He Y 2005 Impact of global warming on water resources in arid area of northwest China *Journal of Mountain Science* **2**(4) 313-18

[2] Feng Q, Cheng G D, Masao M K 2000 Trends of water resource development and utilization in arid north-west China *Environmental Geology* **39**(8) 831-7

[3] Xia J, Zhang L, Liu C, Yu J 2007 Towards better water security in north China *Water Resource Management* **21**(1) 233-47

[4] Shi Y F, Qu Y G 1992 Water resources carrying capacity and its rational utilization in Urumchi River Basin *Beijing: Science Press (in Chinese)*

[5] Xu X Y, Wang H, Gan H 1997 The theory and method of macroeconomic water resources planning in north China Zhengzhou *Yellow River Water Conservancy Press (in Chinese)*

[6] Ruan B Q, Shen J 1997 Calculating model for moderately bearing capacity of regional water resources *Journal of Soil and Water Conservation* **4**(3) 57-61 *(in Chinese)*

[7] Jia R, Xue H F, Xie J C, et al. 1998 Research on regional water resources carrying capacity. Journal of Xian University of Technology **14**(4) 382-7 *(in Chinese)*

[8] Wang J H, Jiang D, Gu D F, et al. 1999 The forecast research of arid city water resources carrying capacity based on SD model. Geography and Territorial Research **15**(2):18-22 *(in Chinese)*

[9] Hui Y H, Jiang X H, Huang Q. 2001 On system dynamic simulation model of water resources bearing capacity in duality mode. Geographical Research **20**(2):191-8 *(in Chinese)*

[10] Zuo Q T 2005 Water resources carrying capacity in urban city: theory, method, application *Beijing: Chemical Industry Press* 22-4 *(in Chinese)*

Authors

Yinhui Long, born in October, 1988, Guangxi Province, China

Current position, grades: assistant engineer.
University studies: Hebei University of Engineering, China.
Scientific interest: Water Resources management.
Publications: 20.

Zhongxiao Guo, born on October 24, 1958, Inner Mongolia Province, China

Current position, grades: professor.
University studies: Hohai University, China.
Scientific interest: Water Resources management.
Publications: 15.

Zilong Liao, born in November, 1987, Guangxi Province, China

Current position, grades: doctoral student.
University studies: China Institute of Hydropower and Water Resources Research.
Scientific interest: hydrology and water resources.
Publications: 10.

Yifan Song, born in January, 1990, Henan Province, China

Current position, grades: postgraduate student.
University studies: China Institute of Hydropower and Water Resources Research.
Scientific interest: hydrology and water resources.
Publications: 3.

The application of saturation memetic algorithm in economic load dispatch

Jianguo Wang[1]*, Cong Cong[2], Xuhua Qing[3], Xiao Han[4]

[1]*School of Automation Engineering, Northeast Dianli University, Jilin, China*

[2]*School of Control and Computer Engineering, North China Electric Power University, Beijing, China*

[3]*Jilin Electric Power Research Institute, Jilin, China*

[4]*College of Information Engineering, Northeast Dianli University, Jilin, China*

Abstract

To solve a nonlinear constrained economic load dispatch (ELD) problem in which the minimized generation cost is taken as objective and valve point effect of thermal units and operation constraints of power grid are taken into account. A saturation memetic algorithm was presented. To avoid the blindness in the search by the proposed using saturation increase efficiency and using adaptive penalty function to deal constraints. Calculation results of 3-machine system verify the effectiveness of the proposed algorithm. Compared with the other existing techniques, the proposed algorithm has been found to perform better. This method seems to be a promising alternative approach for solving the ELD problems in practical power system.

Keywords: memetic algorithm, economic load dispatch, valve point, saturation

1 Introduction

Economic load dispatch (ELD) is an important optimization task in power system operation. The practical ELD problems with valve-point are represented as a non-smooth optimization problem with equality and inequality constraints, and this makes finding the global optimum difficult. Not only traditional mathematical methods are difficult to solve [1,2], but many intelligent optimization algorithms have their own limitations. Such as chaos optimization algorithm (COA) [3,4], the genetic algorithm (GA) [5] artificial immune algorithm (IA) [6], (PSO) [7], the free search algorithm [8], and the differential bee colony algorithm [9].

In this paper, an alternative approach is proposed to the ELD problem using a modified MA. In general, MA is a synergy of evolution and individual learning, which improving the capability of evolutionary algorithms like GA for finding optimal solutions in function optimization problems with accuracy and convergence speed. Genetic Algorithm due to its good exploration capability is used as main algorithm and simulated annealing are used as local searches. And use the variable scales penalty function to improve the performance of the algorithm.

2 Economic load dispatch mathematical model

The ELD problem is to find the optimal combination of power generation that minimizes the total cost while satisfying the total demand. The most simplified cost function of each generator can be represented quadratic function as given in Equation (1) whose solution can be obtained by the conventional mathematical methods 10]:

$$obj = \min \sum_{i=1}^{n} F_i \, , \tag{1}$$

$$F_i = a_i \cdot P_i^2 + b_i \cdot P_i + c_i \, , \tag{2}$$

where *obj* is total generation cost; *a, b, c* is cost coefficients of generator; *n* is number of generators in the system; P_i is output generation for unit *i*.

While minimizing the total generation cost, the total generation should be equal to the total system demand plus the transmission loss. This gives the equality constraint:

a) power balance constraints:

$$\sum_{i=1}^{n} P_i = P_{total} + P_s \, , \tag{3}$$

where P_{total} is the total system demand and P_s is the transmission loss.

b) the generation output of each unit should be between its minimum and maximum limits.

That is, the following inequality constraint for each generator should be satisfied generating capacity constraints:

$$P_{i,\min} \leq P_i \leq P_{i,\max} \, , \tag{4}$$

where $P_{i,min}$, $P_{i,max}$ is the minimum, maximum output of generator.

Corresponding author e-mail: darcertheword@126.com

In the actual operation, the turbine intake valve suddenly turned caramelized phenomenon consumption characteristics of the unit will generate valve point effect. To model the effects of valve-points, a recurring rectified sinusoid contribution was added to the input-output Equation [11,12]. The result is:

$$E_i = | e_i \sin(f_i(P_i - P_{i,min})) | . \qquad (5)$$

The fuel-cost function considering valve-point loadings of the generating units is given as:

$$F_i = a_i \cdot P_i^2 + b_i \cdot P_i + c_i + E_i , \qquad (6)$$

Transmission losses are a function of the unit generations and are based on the system topology. Solving the ELD equations for a specified system load requires an iterative approach since all unit generation allocations are embedded in the equation for each unit.

In practice, the loss penalty factors can be calculated directly using the B matrix loss formula. B coefficients are used for this example. The relations between transmission losses and Generator active power:

$$P_s = P^T B P + P^T B_0 + B_{00} , \qquad (7)$$

where P is the N-dimensional generator active power vector; B is loss coefficient; the B_0 dimensional symmetric matrix for N-dimensional vector, B_{00} is constant.

3 Memetic-based ELD

Memetic algorithm [15, 16] is a hybrid algorithm population-based global search and individual-based local search. Algorithm is a framework, using different search strategies can constitute different algorithms. For solving specific problems, use genetic algorithms as the global search method, use simulated annealing algorithm as the local search strategy.

The unique aspect of MA is that the chromosomes are facilitated to gain some experience with a local search process in between regular evolutionary process. Similar to the GA, MA also generates an initial population randomly and searches in the fitness landscape. Subsequently, the local search process leads solutions in the direction of local optima. These improvements are going to accumulate over all generations, resulting in a larger improvement in total performance.

In order to avoid the blindness of search, introduced the concept of saturation, do a selective mutation, crossover and selection. Saturation is a metric which set on the solution space. For minimum optimizing, the more smaller value and saturation, the more close to the optimal solution. On the contrary, far from the optimal solution. In searching process, through saturation adjusts populations. High saturation means small choice probability, so as to ensure the diversity of population. Always searching for the optimal solution from the global, ensures that focus on the local optimal solution, while maintaining the global optimal solution.

$$s(x) = \frac{1}{1 + q(x)} , \qquad (8)$$

$$q(x) = \frac{f - f_{min}}{f_{max} - f_{min}} , \qquad (9)$$

where $s(x)$ is saturation F_{max}, F_{min} is maximum and minium respectively, F_i is the objective $q(x)$ is intermediate amount.

For every violated constraint a penalty term proportional to the amount of the constraint violation is added to the fitness function. The penalty factor is chosen sufficiently large to discourage the selection of solutions with violated constraints.

The penalty function PF is:

$$PF = MAX (0, | P_{Total} - \sum_{i=1}^{n} P_i | \times \lambda(t)) , \qquad (10)$$

where $\lambda(t)$ is penalty factor. In this paper using adaptive penalty function for constraints, according to the optimal solution meet the degree of constraints adjust the penalty function. Adaptive penalty function is constructed as follows:

$$\lambda(t+1) = \begin{cases} \lambda(t) / (2\beta_1 / (\beta_1 + \beta_2)), & case1 \\ \lambda(t) / (2\beta_2 / (\beta_1 + \beta_2)), & case2 \\ \lambda(t), & case3 \end{cases} \qquad (11)$$

where β_1 is a random number between $(1,2)$, β_2 is a random number between $(0,1)$; $case1$ reduce the penalty factor to increase Infeasible solution; $case2$ increase the penalty factor to increase the feasible solution; $case3$ the penalty factor remain unchanged. Fitness function becomes:

$$\min F = \min(\sum_{i=1}^{n} F_i + PF) , \qquad (12)$$

The hyper-heuristic search mechanism in the memetic algorithms offers the speed and quality of convergence. This paper uses Genetic Algorithm as main algorithm and the SA are used as the local search, where both the local search algorithms are heuristic algorithms. The algorithm steps are as follows:

Step1: encoding.

For the application of MA to the ELD problem, a float encoding was chosen to encode a solution. Individual's position at iteration 0 can be represented as real power outputs. Initialization of a group at random while satisfying constraints.

Step2: fitness functions.

Evaluation of the chromosome string is calculating the objective function for the problem.

Step3: genetic operators.

a) *select*. This operator use roulette, individuals which high fitness value have a greater chance to be selected to the next round of operations.

b) *crossover*. This operator is applied with a certain probability. When applied, the parent genotypes are combined (exchange bits) to form two new genotypes that inherit solution characteristics from both parents. In the opposite case the off-springs are identical replications of their parents.

$$\begin{cases} P_1^{new} = \omega_1 P_1 + (1-\omega_1)\cdot P_2 \\ P_2^{new} = \omega_2 P_2 + (1-\omega_2)\cdot P_1 \end{cases}. \qquad (13)$$

c) *mutation*. With a small probability, randomly chosen bits of the offspring genotypes change.

$$P = P + (b_{sub} - P)\cdot [r\cdot(1-t)]^2, \qquad (14)$$

Variable step size crossover and mutation probability

$$\begin{cases} P_c^{j} = P_c^{j-1} - (P_c^{o} - 0.6)/\max iter \\ P_m^{j} = P_m^{j-1} + (0.1 - P_m^{0})/\max iter \end{cases}, \qquad (15)$$

d) *elitism*. The best solution of every generation is copied to the next so that the possibility of its destruction through a genetic operator is eliminated.

If any element of an individual violates its inequality constraint then the position of the individual is fixed to its maximum/ minimum operating point.

Step4: local search strategy

Local search strategy based on simulated annealing.

Step5: stopping criteria:

The MA is terminated if the iteration approaches to the predefined maximum iteration.

4 Examples and results

Proposed MA algorithm has been applied to ELD problems in three different test cases which the objective functions can be either smooth or non-smooth. The results obtained from the MA are compared with those of other methods: chaos optimization algorithm (COA) [3], the genetic algorithm (GA) [5], artificial immune algorithm (IA) [6], (PSO) [7], the free search algorithm [8].

Three separate runs were made to demonstrate the ability of the program using different initial random seed numbers. The cases considered were:

1) Input-output curves neglecting Valve-point effect and transmission losses.

2) Input-output curves considering Valve-point effect neglecting transmission losses.

3) Input-output curves considering Valve-point effect and transmission losses.

TABLE 1 Units data for test case

No.	*a*	*b*	*c*	*e*	*F*	P_{min}	P_{max}
1	0.00156	7.92	561	300	0.0315	100	600
2	0.00194	7.85	310	200	0.0420	100	400
3	0.00482	7.97	78	150	0.0630	50	200

The transmission loss coefficients (*B* coefficients) used for the power system follow:

$$B = \begin{bmatrix} 0.676\times10^{-3} & 0.953\times10^{-4} & -0.507\times10^{-4} \\ 0.953\times10^{-4} & 0.521\times10^{-3} & 0.901\times10^{-4} \\ -0.507\times10^{-4} & 0.901\times10^{-4} & 0.294\times10^{-3} \end{bmatrix},$$

$$B_0 = \begin{bmatrix} -7.66\times10^{-2} & -0.342\times10^{-2} & 1.89\times10^{-2} \end{bmatrix}^T,$$

$$B_{00} = 4.0357.$$

To compare the results between MA and various methods in a statistical manner, every case was calculated 50 times. All the programs were run on a 1.8 GHz, with 2GB RAM PC. The crossover and mutation rates for binary GA were chosen as 60% and 0.05%, respectively, following common literature. The optimal result is provided.

The input data of the system are given in [3], where the system demand considered is 500MW. Table 2 shows the comparison of the results from MA, COA [3], GA [5], IA [6] and IFS [8]; Table 3 shows the results of considering valve-point effect neglecting transmission losses; Table 4 shows the results of considering valve-point effect and transmission losses.

TABLE 2 Comparison among different methods neglecting valve-point effects (demand= 500 [MW])

Method	P_1	P_2	P_3	∑P	Cost
GA	214.60	227.20	58.20	500	5084.31
IA	237.73	187.58	74.69	500	5083.04
COA	228.71	202.29	69.01	500	5082.33
IFS	228.93	202.15	68.92	500	5082.22
MA	229.00	202.13	68.87	500	5082.22

TABLE 3 The results of considering valve-point effect neglecting transmission losses (demand= 500 [MW])

Method	P_1	P_2	P_3	∑P	Cost
GA	299.22	101.34	99.30	499.86	5131.74
IA	199.76	250.24	50.00	500.00	5121.47
COA	299.41	100.70	99.90	500.01	5095.47
IFS	199.74	250.25	50.01	500.00	5095.40
MA	199.73	250.27	50.00	500.00	5095.38

TABLE 4 The results of considering valve-point effect and transmission losses (demand= 500 [MW])

Method	P_1	P_2	P_3	∑P	Ps	Cost
GA	295.09	173.43	101.71	570.23	70.23	5772.27
IA	299.41	172.00	98.84	570.30	70.24	5735.93
COA	299.46	171.95	98.82	571.23	71.23	5735.76
IFS	299.45	171.91	98.85	571.21	71.21	5735.74
CPSO	299.47	171.88	99.87	571.22	71.22	5735.89
MA	299.47	171.91	99.86	571.22	71.22	5735.72

As seen in Tables 2, 3 and 4, the MA has provided the global solution with a very high probability, exactly satisfying the equality and inequality constraints.

Whether to consider the effect of transmission losses and valve point, the result is difficult, in the actual should consider the valve point effect and transmission loss. The MA has provided the global solution satisfying the constraints with a very high probability for the ELD

problems with smooth cost functions. For the ELD problems with non-smooth cost functions due to the valve-point effects, the MA has also provided the global solution with a high probability for 3-generator system which are better than other methods. The MA has shown superiority to the conventional numerical method.

method performs good convergence property and can avoid the premature convergence as compared to other numerical method and artificial intelligence algorithm to obtain better quality solution. This promises a great potential of the proposed method for real-time economic load dispatch.

5 Illustrations

This improved memetic algorithm has been successfully implemented to solve ELD problems with the generator constraints. Many nonlinear characteristics of the generator have been considered. The proposed MA

Acknowledgments

References

[1] Peng L 2008 *Southern Power System Technology* **2**(5) 43-6
[2] Shushan L, Gang L, Chuntian,C 2011 *Proceedings of the Csee* **31**(7) 41-7
[3] We T, Dianpu T 2000 *Proceedings of the Csee* **20**(10) 36-40
[4] Fang Shuangxin W, Hengjun H 2005 *Proceedings of the Csee* **25**(24) 90-5
[5] Walters D C, Sheble G B 1993 *IEEE Transactions on Power Systems* **18**(3) 1325-32
[6] Jiaju Wenchuan M, Xiaomeng Q B 2006 *Power System Technology* **30**(23) 41-4
[7] Wenchuan M, Jiaju Q 2007 *Proceedings of the Chinese Society of Universities for Electric Power System and Automation* **19**(2) 114-9
[8] Shu X, Jianhua Z, Wei Q 2010 *Power System and Clean Energy* **26**(3) 37-40
[9] Liwei X, Mai X, Baoguo X 2011 *Control and Decision* **26**(12) 1813-7
[10] Jianguo W, Rudong M 2012 T*hermal Power Generation* **41**(12) 70-3

[11] Xiaohui Y, Yanbin Y, Cheng W 2005 *Transactions Of China Electro Technical Society* **20**(6) 92-6
[12] Zhiqiang Y, Zhijian H, Chuanwen J 2002 *Proceedings Power Con, International Conference,IEEE, Kunming, China* 2313-7
[13] Yalin M, Guozhong Z, Bin Z 2005 *Proceedings of the CSEE* **25**(3) 65-70
[14] Yunhe H, Xingeng X, Yaowu W 2003 *Proceedings of the CSEE* **23**(3) 59-64
[15] Hongfeng W, Dingwei W, Min,H 2010 *Control Theory & Applications* **27**(8) 1060-8
[16] Zhihao W, Xiangwei Z, Hongwei M 2013 *Journal of Chinese Computer Systems* **34**(3) 617-20
[17] Zhuo L, Chun H, Zhenhua G 2011 *Power System Technology* **35**(2) 100-4
[18] Weiping D, Jiandong, W Weihua 2011 *Journal of Shandong University(Natural Science)* **46**(5) 97-102

Authors

Jianguo Wang, born in October, 1963, Jinlin, China

Current position: professot in Northeast Dianli University.
University studies: Department of automation in Tientsin University.
Scientific interest: intelligent instrument.
Publications: 2 patents, 50 papers.

Cong Cong, born in August, 1987, Liaoning, China

Current position: Ph.D student in North China Electric Power University.
University studies: Control Theory and Control Engineering in North China Electric Power University.
Scientific interest: state estimation.
Publications: 3 papers.

Xuhua Qin, born in February, 1962, Jinlin, China

Current position: engineer in Jilin Electric Power Research Institute.
University studies: Department of automation in Tianjin University.
Scientific interest: smart grid technology.

Xiao Han, born in November, 1986, Liaoning, China

Current position: Master student in Northeast Dianli University.
University studies: Computer Science and Technology in Northeast Dianli University.
Scientific interest: intelligent information processing.
Publications: 4 papers.

Modelling and hull vibration calculation of very large container ship

Fenglei Han*, Ankang Hu, Yachong Liu, Chunhui Wang

Harbin Engineering University, Nantong str. 145, Harbin, P.R. China

Abstract

Hull vibration is inevitable in large vessel's operation. Too large hull vibration will not only cause damage to the hull structure, but also will affect the normal use of marine equipment and crew at work and life. Therefore, to predict the performance of the ship's structure vibration in ship design stage is essential, which can guide reasonable solutions and necessary damping measures. The vibration calculation report includes: Ship's free vibration calculation and forced vibration response calculations caused by vibration source on board the ship. The former is mainly to obtain the free vibration characteristics of the whole ship structure, to find hull's 1~3 order of vibration modes which are most likely to occur and the corresponding frequency values. The latter is mainly get the vibration response of hull ship's forced vibration caused by excitation source, to avoid low-frequency vibration's affect to the work and life of the crew. Its response should meet the relevant standards and requirements.

Keywords: modelling, free vibration, forced vibration, container ship, FEM

1 Introduction

The rapid increase in speed and size of the ships constructed during the 1960's led to the realization that wave-induced ship hull vibrations can give rise to significant stresses in the hull. The vibrations are usually classified as either whipping or springing, depending on whether the vibration mode is transient or steady state [1]. Ship hull vibration can be generally classified into two categories, global and local vibrations. For global vibration, the whole hull girder of a ship is vibrating in response to the excitation at particular revolutions of the main engines, propellers and auxiliary machinery, or from water waves [2]. Local vibration occurs when only an isolated part of the ship structure is in resonance [2,3]. Local resonance can be treated locally by modifying the resonating structural component or by adding vibration absorption and damping devices. Nevertheless, excessive deformation of the ship hull is more likely to come from global vibration, particularly at the first few fundamental modes of the ship structure.

Generally speaking, the global vibration of a ship, including natural frequencies and mode shapes of a complete ship hull structure, is analysed by strip theory where natural frequencies of the entire ship are calculated from beam theory and the estimation of weight and moment distributions at each strip of the ship hull. A method of such an approach is given by Todd [2]. Van Gunsteren also investigated wave-induced ship hull vibration [4], the two-node vertical vibration mode of two ship models using a modified strip theory is calculated. Because of the fast advance of computer technology and

the increasing speed and capacity of modern computers, it's possible to analyse the low frequency dynamic response of a complete ship structure in a three-dimensional model using finite element analysis (FEA). FEA is used increasingly in the analysis and design of complex ship structures [5].

Lin and Pan [6, 7] used a closed form solution to define the characteristics of input mobility [8] of finite ribbed plate structures to force and moment excitations. Their research showed that the point force input mobility of a finite rib-stiffened plate is bounded by those of the corresponding unribbed plate and beam(s) of the ribbed plate. The input mobility is dominated mainly by the beam bending stiffness when the force excitation is applied to the beam, and its plate stiffness controlled when the beam is more than a quarter of plate bending wave length away from the force location. The result has been verified experimentally by Nightingale and Bosmans [9]. The torsional moment input mobility of a ribbed plate (for ribs having relatively small torsional stiffness) is dominated by the plate bending stiffness and its frequency-averaged value can be represented by that of the corresponding infinite plate. Lin [7, 10] extended the study to further investigate the characteristics of wave propagation and attenuation of finite periodic and irregular ribbed plates by employing a modal approach. He found that vibration of a ribbed plate structure can be confined by imposing irregularity to rib locations on the plate. The study of vibration characteristics of rib-stiffened plates provides a general understanding of wave propagation and its control mechanism in ribbed plate structures. However, when the structure becomes complex as in the case of a complete

---

ship hull structure, the complexity of analysis increases due to the coupling of different wave types and multiple wave propagation paths in the structure[11].

In this paper, the calculation includes two parts: Calculation and analysis of the hull free vibration and forced vibration caused by vibration source on board. The former is mainly to ascertain the free vibration characteristics of the hull girder to determine whether the hull girder may be related to the main exciting force resonance under normal operating conditions. The latter is to ascertain the forced vibration characteristics (velocity and acceleration) of hull girder and superstructure caused by hull excitation source under normal operating conditions. To compare the vibration response values of some important local structure in the superstructure and engine room with the corresponding requirements of ISO standards to check whether necessary measures to be taken to improve the vibration performance of the ship [12].

2 Large container ship's finite element model

The calculation is based on the 9200TEU container ship designed by CIMC Naval Architecture and Ocean Engineering Design Institute. The finite element model used in 9200TEU container ship's vibration calculation is built entirely in accordance with the relevant design drawings, and the processing and analysis operations is completed by commercial finite element analysis software MSC/PATRAN and NASTRAN.

Global coordinate system of the model is right hand Cartesian coordinate system:

- X direction goes along ship's length pointing to bow;
- Y direction goes along ship's breadth pointing to Portside;
- Z direction goes along ship's depth pointing to deck.

Structural model and applied loads are in International System of Units (N, mm, s)

All plate structures, such as shell, transverse bulkhead, inner bottom, web frame and longitudinal bulkheads etc. are modelled by CQUAD4 and CTRIA3 shell elements. All girders and stiffeners are modelled by eccentric beam with appropriate combination. Small structures with hole are adjusted when finally balancing the quality of the whole ship structure and large structures with holes are modelled according to their actual shape as possible.

Mesh size in longitudinal direction is strong frame spacing and in transverse direction is longitudinal spacing. There are 2 elements in vertical direction between platforms. The floors and girders in double bottom are divided into 3 elements along the height direction.

Finally, there are and 450000 finite elements and 1320000 degree of freedom. Finite element models used in different loading conditions are shown in Figures 1 and 2.

During the process of establishing the finite element model the hull weight and the position of gravity centre are controlled by adjusting density of material of some elements and adding structural mass point. The deadweight in different loading conditions is adjusted by adding structural mass points at corresponding loading position and the mass points are linked to surrounding nodes by MPC. Final total weight and gravity centre of the model and those in loading manual are shown in Table 1.

Hull added mass is computed by empirical equation. Vertical and horizontal added mass is respectively computed in each loading condition. The computed added mass is added to shell underwater in the form of concentrated mass point.

In the computation of vertical vibration, the added mass is computed by the formula proposed by F M Lewis and F H Todd:

$$m_{av} = \frac{1}{2}\pi a_v C_V K_i \rho b^2 , \tag{1}$$

added mass formula in horizontal vibration is:

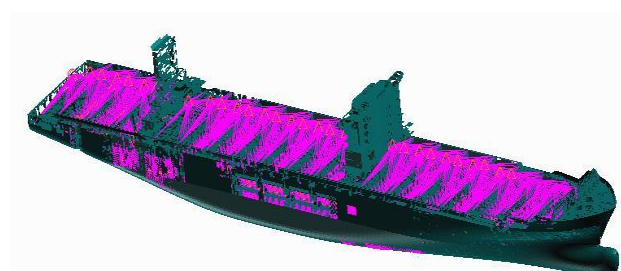$$m_{aH} = \frac{1}{2}\pi a_H C_H K_i \rho d^2 . \tag{2}$$

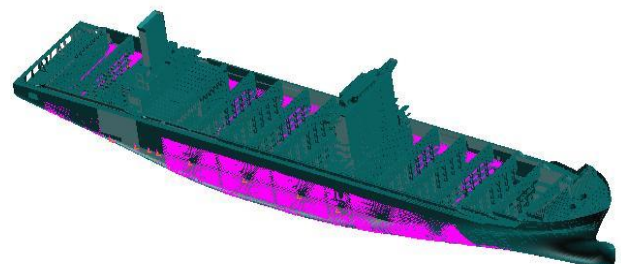FIGURE 1 14TD loading condition element modal

FIGURE 2 Ballast loading condition element modal

TABLE 1 Finite element modal's weight adjustment

Loading Condition	Ballast Arrival		14TD	
	Loading manual	Finite element model	Loading manual	Finite element model
Mass	66958.4	66100	147442.0	146900
Centre of gravity-X	-16.09	-17	-6.13	-6.4
Centre of gravity-Y	0	0.073	0	9.09e-2
Centre of gravity-Z	11.48	11.48	18.676	20.5

3 Free vibration analysis and calculation of global structure

Calculation and analysis of free vibration of global ship structure is to test whether the main excitation source frequency is close to the natural frequency of the hull, so as to modify the design to avoid resonance when necessary. In global ship structure analysis of free vibration, a very important part is to simulate the ship free floating in the water without boundary conditions.

Table 2 shows the typical free vibration frequencies from FEM calculation, some of the global ship typical free vibration typical modes under the 14TD loading conditions are shown in Figure 3-5, and Figure 6-8 represent the global ship typical free vibration modes under the ballast loading conditions.

Main engine excitation frequency is:

$$f = \frac{78}{60} = 1.3 Hz , \qquad (3)$$

which is close to the 2-nd order vertical natural vibration, frequency1.356Hz under the BLSA condition, the difference is 4.3%; compared with the 3 order vertical natural vibration frequency under the 14TD condition is 1.476Hz, the difference is 13.53%, satisfies the requirements.

According to the "Ship vibration control guide 2000", in order to avoid the resonance between the low order vibration frequencies and the excitation frequencies, the 1-3 order natural frequencies must be different from the excitation frequency at the level ±8%～±10%, ±10%～±12%, ±12%～±15%.

If the requirements are not satisfied, the vibration response calculation and the vibration response measurement are necessary, which will be done according to the ISO standard (6954 resolution, 1979.9): vertical and horizontal vibration evaluation criteria of merchant ships, whose length are greater than 100 m. If the response values are greater than the standard values, vibration reducing measures should be taken.

TABLE 2 Typical free vibration frequencies from FEM calculation (Hz)

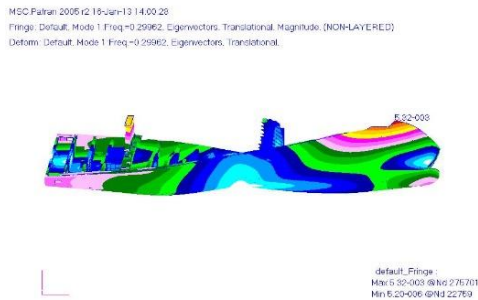| Loading condition	Vertical vibration			Torsional vibration	Horizontal& Torsional vibration				
	1 order	*2 order*	*3 order*	*1 order*	*1&2 order*	*2&2 order*	*2&3 order*	*3&3 order*	*3&4 order*
14TD	0.470	0.897	1.476	0.354	0.603	0.933	1.056	1.279	--
BLSA	0.716	1.356	1.914	0.516	0.674	1.43	1.68	--	--

FIGURE 3 1 order torsional vibration mode under the14TD loading condition (0.299Hz)
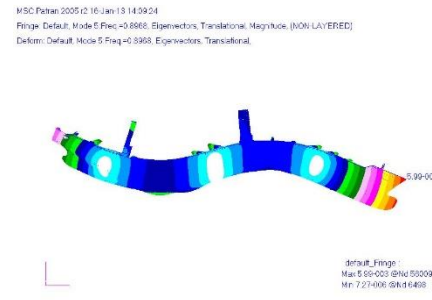

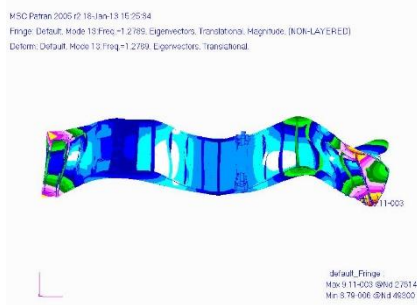FIGURE 4 2 order vertical vibration mode under 14TD loading condition (0.897Hz)

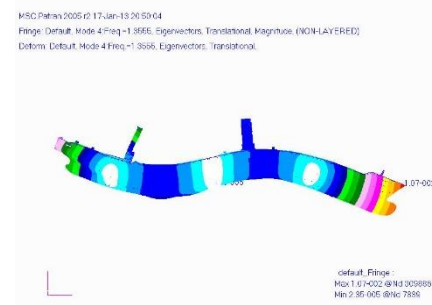FIGURE 5 3 order horizontal and torsional vibration modes under 14TD loading condition (1.279Hz)

FIGURE 6 2 order vertical vibration mode under BLSA arrival loading condition (1.356Hz)

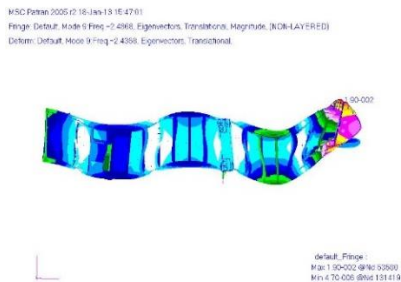Han Fenglei, Hu Ankang, Liu Yachong, Wang Chunhui

FIGURE 7 3 order horizontal and torsional vibration modes under BLSA arrival loading condition (2.437Hz)
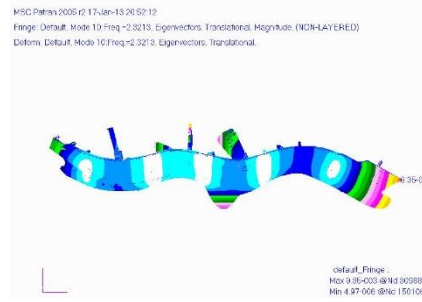

FIGURE 8 4 order vertical vibration mode under BLSA arrival loading condition (2.32Hz)

4 Global ship forced vibration response calculation

A Global ship forced vibration frequency response analysis considering the excitation force, which is the main excitation force, is a method to define the whole hull vibration response characteristics, so as to make an investigation of the working and living comfort on the supper structure.

Theoretically, propeller excitation force is the surface force, which is calculated from the integral along the hull surface, the force applying area is D×D (D is the diameter of the propeller). When the bottom is V type, the excitation force distribution is very irregular, accursedly usually, the value calculated by the propeller surface pressure formulation is used in the dynamic response calculation. The pressure could be applied on the active area directly. In this calculation the measured values of the typical points provided by the manufacturer is used, the applied area is D×D, about 8.9m*8.9m. Figures 9 and 10 are excitation force diagrams.
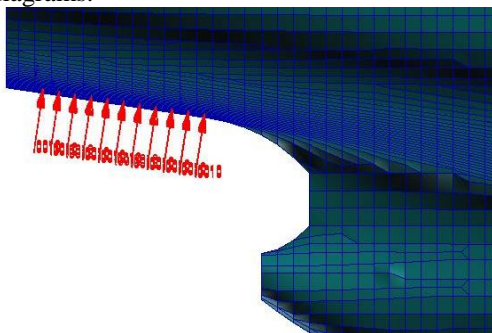

FIGURE 9 Propeller excitation force acting diagram (14TD loading condition, draft 14.8m)
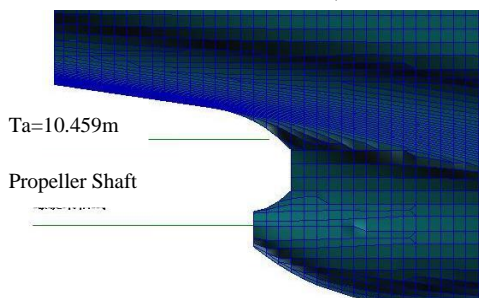

FIGURE 10 Excitation force range diagram (BLSA arrival loading condition, there is almost no pressure on the bottom plate)

Basically the structural damping could be obtained from ship experiment. However, in calculation usually the advantage of the damping coefficients obtained from classification societies through a large number of experiments, such as DNV and GL defined the damping coefficients in vibration response calculation, could be used. In this report we use the damping coefficient defined by GL in the vibration response calculation of 2750TEU container ship, the damping coefficient is 1% when the frequency is 0Hz, 8% when the frequency is 20Hz; under the BLSA loading conditions that is 0.5% when the frequency is 0Hz, 6% when frequency is 20Hz, shown as in Figure 11.
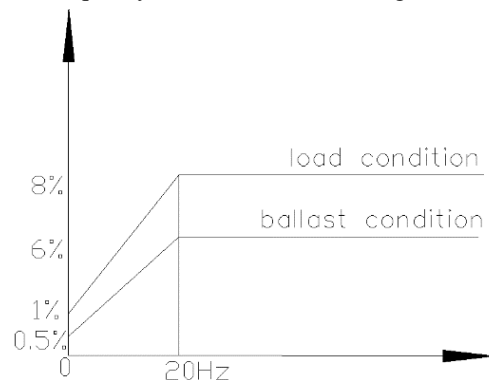

FIGURE 11 Damping factor to frequencies

Main engine excitation source has been shown in Table 3, the vibration excitation source would be loaded on the centre of gravity of the main engine through the junction points of the main engine and the internal bottom plate, the local structure diagram is shown in Figure 12.
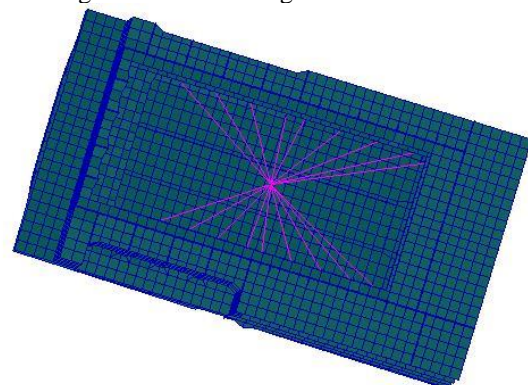

FIGURE 12 Local structure diagram with excitation source of main engine

TABLE 3 Engine imbalance force and imbalance moment

External forces(KN)		Guide force X-moments in (KNm)	
1.order:Horizontal	16	1. order:	503
1.order:Vertical	16	2. order:	21
2.order:Vertical	19	3. order:	1905
4.order:Vertical	80	4. order:	1947
6.order:Vertical	3	5. order:	1987
External moments(KNm)		6. order:	4103
1.order:Horizontal a)	835	7. order:	758
1.order:Vertical a)	835	8. order:	545
2.order:Vertical	764	9. order:	100
4.order:Vertical	468	10. order:	176
6.order:Vertical	107	11. order:	126
Guide force H-moments in (KNm)		12. order:	174
1x No.of cyl.	1078	13. order:	61
2x No.of cyl.	97	14. order:	76
3x No.of cyl.	-	15. order:	140
		16. order:	76

The velocity and acceleration to frequencies curves of the typical points are shown in the Figures 13-20; the absolute values of velocities and accelerations are used.
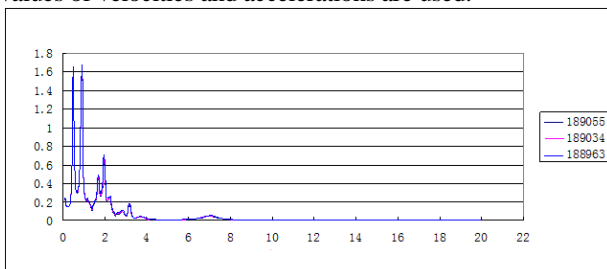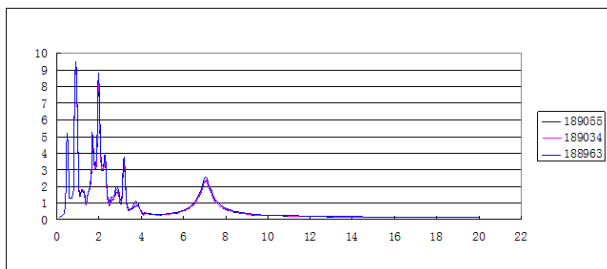

FIGURE 13 Velocity curves of typical points on A deck

FIGURE 14 Acceleration curves of typical points on A deck
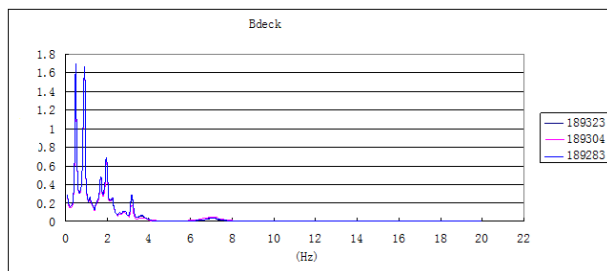

FIGURE 15 Velocity curves of typical points on B deck
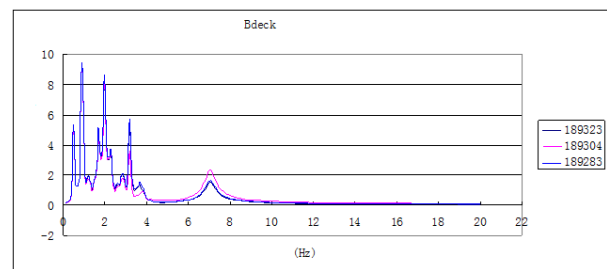

FIGURE 16 Acceleration curves of typical points on B deck
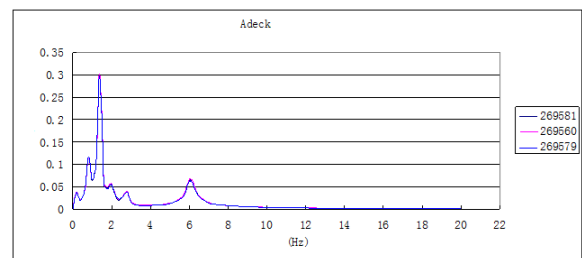
FIGURE 17 Velocity curves of typical points on A deck
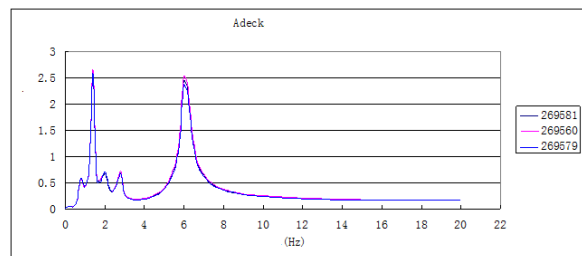
FIGURE 18 Acceleration curves of typical points on A deck
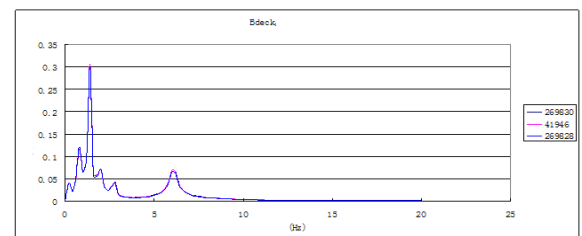
FIGURE 19 Velocity curves of typical points on B deck
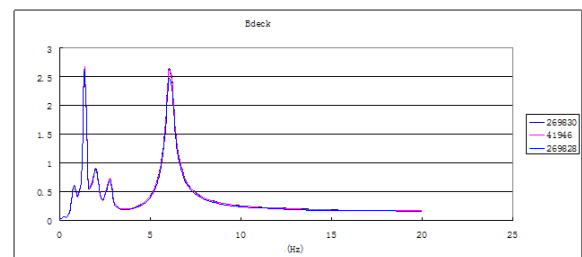
FIGURE 20 Acceleration curves of typical points on B deck

By the velocity and frequency response curves and acceleration frequency response curves it could be shown that, when the frequency is equal to 0.498Hz and 0.896Hz, velocity and acceleration response become larger values. Compared to the free vibration mode of 14TD loading condition, 0.47Hz is the frequency of the 1 order vertical vibration mode. So when the excitation force frequency is 0.498Hz, the vibration response of the local structure of the supper structure will be larger; 0.897Hz is the natural frequency of 2 order vertical vibration, so when the excitation force frequency is 0.897Hz the second vibration peak value appears. However all the calculated velocities and accelerations are under the corresponding criteria values in ISO6954 (2000E).

From the velocity and acceleration frequency response curves it could be seen that, with the increase of vibration frequency, vibration response amplitude increases as well, when the excitation frequency approach to the 1st order natural frequency of the hull girder. The amplitude will reach the maximum value firstly, at this moment the mode of the

forced vibration is close to the 1 order mode, and the resonance will occur.

When the excitation frequency is further increasing, the hull response decreases quickly, then increases gradually, until the excitation frequency approach to 2 order natural frequency, the amplitude will reach the second peak point, and the second resonance is going to happen, followed by the third order, fourth order resonance and so on.

Through calculation and analysis, under this condition, the velocity response and acceleration response of typical points satisfy the corresponding requirements.

5 Conclusions

The calculation of free vibration and forced vibration of 9200TEU container vessel is carried on in a 3D FEM environment, with the help of the large commercial FEM software MSC/PATRAN and NASTRAN to achieve, by adjusting the material density and applying concentrated

mass to equal the loading mass distribution in the manual. Added mass is calculated by empirical formula calculate, and added to the mass points below the water line.

From the calculation, we can see that, vibration response amplitude increases with the increase of vibration frequency, the amplitude will reach the maximum value when the excitation frequency approach to the 1st order natural frequency of the hull girder, the resonance will occur at the moment the mode of the forced vibration is close to the 1 order mode. And hull response decreases quickly when the excitation frequency is further increasing, then increases gradually, until the excitation frequency approach to 2 order natural frequency, the amplitude will reach the second peak point, and the second resonance is going to happen, followed by the third order, fourth order resonance etc.

Through the analysis of free vibration and forced vibration of 9200TEU container ship, most of the forced vibration response values of checking points are less than the standard criteria and satisfy the vibration requirements.

References

[1] Jensen J J 1996 Wave-induced Ship Hull Vibrations in Stochastic Seaways, Marine Structures **9**(3) 353-4
[2] Todd F H 1961. *Ship hull vibration*. London: Edward Arnold Ltd.
[3] Ward F, Norris C, Catley D, Crexis A 1982 Local vibrations in ship's structures. *Trans North East Coast Inst Eng Shipbuilders* **98** 49-64
[4] Van Gunsteren F F 1974 Some further calculations of wave-induced ship hull vibrations *Proceedings of Dynamics of Marine Vehicles and Structures in Waves* New York Mechanical Engineering Publications 90
[5] Hughes O F 1988 Ship structural design Jersey city New Jersey *Society of Naval Architects and Marine Engineers* 1988
[6] Lin T R, Pan J 2006 A closed form solution for the dynamic response of finite ribbed plates *J Acoust Soc Am* **119**(2) 917
[7] Lin T R 2006 Vibration of finite coupled structures, with applications to ship structures *PhD dissertation The University of Western Australia*

[8] Fahy F J, Gardonio P 2007 Sound and structural vibration radiation, transmission and response 2nd ed *Amsterdam Academic Press*
[9] Nightingale T R T, Bosmans I 2006 On a drive point mobility of a periodic rib-stiffened plate *Proceedings of Inter-Noise* Hawaii Washington D C INCE-USA 3364-73
[10] Lin T R 2008 A study of modal characteristics and the control mechanism of finite periodic and irregular ribbed plates *J Acoust Soc Am* **123**(2) 729-37
[11] Lin T R, Pan J, O'Shea P J, Mechefske C K 2009 A study of vibration and vibration control of ship structures *Marine Structures* **22**(4) 730-43
[12] Jin X, Xia L 2011 Hull vibration *Shanghai Jiao Tong University Press* Shanghai

Authors

Fenglei Han, born in March, 1980, Heilongjiang, China

Current position, grades: PhD candidate at Harbin Engineering University. Research assistant of Harbin Engineering University.
University studies: Master of Engineering, Aircraft Design, at Bauman Moscow State Technical University, Russia, in 2010.
Scientific interest: Ship structural mechanics.
Publications: 2 Patents, 3 Papers.

Ankang Hu, born in March, 1956, Sichuan, China

Current position, grades: professor of Harbin Engineering University and CEO of CIMC.
University studies: ship building and ocean architecture design in Harbin Engineering University.
Scientific interest: ship building and ocean architecture design.
Publications: 20 Patents, 35 Papers.

Yachong Liu, born in May, 1988, Shandong, China

Current position, grades: PhD student at Harbin Engineering University, China.
University studies: Master of Engineering at Harbin Engineering University, ship building and ocean architecture design.
Scientific interest: ship structural mechanics.
Publications: 3 Papers.

Chunhui Wang, born on March 24, 1985, Henan, China

Current position, grades: PhD candidate at Harbin Engineering University. Research Assistant of Harbin Engineering University.
University studies: Master of Engineering, Aircraft Design, at Bauman Moscow State Technical University, Russia, in 2011.
Scientific interest: Ocean architecture design.
Publications: 2 Patents, 3 Papers.

Experiment of energy recovery efficiency and simulation research on EV's regenerative braking system

Xiaobin Ning[1]*, Ning Li[2], Junping Jiang[3]

[1]*Institute of Vehicle Engineering, Zhejiang University of Technology, China*

[2]*Zhijiang College of Zhejiang University of Technology, Zhejiang University of Technology, China*

[3]*Zhejiang Geely Automobile Research Institute CO LTD, China*

Abstract

The proposed technique ECPS aims to design and implement a constant pressure hydraulic regenerative braking system with flywheel that can operate the problems of short driving range and improve efficiency of braking energy recovery about electrical vehicle. We established ECPS after comparing several hydraulic regenerative braking schemes .Then we explored the correlation between displacement of pump/motor and efficiency of energy recovery by undertook experiment on using variable displacement of pump/motor. After that, we investigated braking and ECE-15 driving condition simulation based on AMESim, evaluated correlation between displacement of pump/motor and efficiency of energy recovery. The results indicate that the driving range of electrical vehicle has been increased by 25% and the service life of battery was prolonged because depth of discharge was decreased.

Keywords: recovery efficiency, regenerative braking, experiment

1 Introduction

Regenerative braking system refers to the vehicle's braking energy recycling; there are a variety of forms, such as Electrical energy recovery and hydraulic energy recovery. Using motor regenerative braking energy recovery systems have a significant effect in pure electric vehicles, but there are still some drawbacks: it limited capacity of the braking energy recovery because of the electric energy storage's energy density is low, thus affecting the driving range of electric vehicles mileage; motor regenerative braking belongs to the electronic system, due to the brake factors, mechanical power system is more reliability than the electric power system [1]. Owing to hydraulic energy recovery power density is higher than the electric energy recovery [2], using hydraulic brake energy recovery in pure electric vehicle under the same conditions and at the same time can able to recover and release more energy, so it can better improve driving range of electric vehicles. In addition, automotive powertrain changes for the use of hydraulic energy recovery is little, relative to the electric energy recovery control link is simple and more reliability.

At present, there are many different forms of hydraulic fuel-efficient vehicles in foreign research and development, depending on its configuration and powertrain combinations in different ways, it can be divided into: series structure, a typical representative of the U.S. EPA's hydraulic hybrid system [3], Its structure is simple, the system is easy to control different parameters; parallel structure, such as Cumulo driving system from Swedish Volvo [4], it changes little on vehicle and has high energy efficiency; hybrid structure, such as the last century nineties developed Constant Pressure Source (referred to CPS) hydraulic drive system in Japan, it has become one of the main forms of vehicle energy recovery system due to its good energy saving and simple structure [5]. CPS system through the engine and flywheel hybrid drive for the power system, the use of Constant Pressure Source hydraulic system to transform energy.

The system in this article will use the rear-mounted parallel structure; CPS flywheel hydraulic regenerative braking system was put forward. Its high energy efficiency and control link is relatively simple, small changes to vehicle power system, also the cost is relatively low.

2 Electric vehicle hydraulic regenerative braking system

According to the CPS, we proposed electric vehicle regenerative braking hydraulic system (referred ECPS), in parallel with the power system, determine the drive system to be rear driveline parallel system. The ECPS system structure shown in Figure 1.

Corresponding author e-mail: nxb@zjut.edu.cn
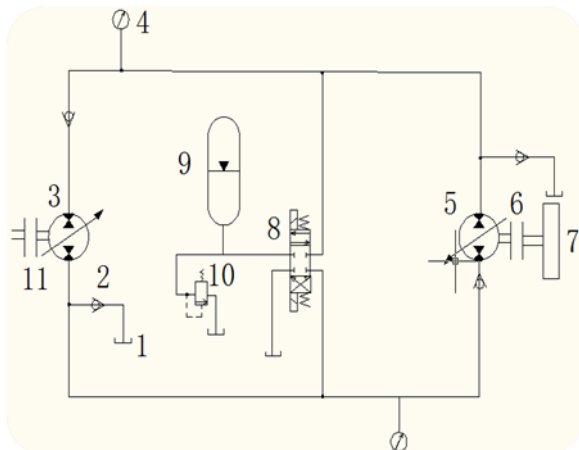
1. Tank 2. Check valve 3. Variable pump / motor connected with the drive axle 4. Manometer 5. Variable pump / motor connected with the flywheel 6., 11. Clutch 7. Flywheel 8. Three-way valve 9. Accumulator 10. Pressure relief valve

FIGURE 1 ECPS system structure

When the car braking deceleration, the drive wheel connected to the 3 pump/motor and the 3 pump/motor works as a pump, the resistance generated by this pump to brake wheel while the transform brake mechanical energy into hydraulic energy, the braking energy recovery. The braking force adjust by the displacement of 3 pump/motor to achieve. High pressure oil flow from 3 to 5, the 6 clutch connected flywheel 7 with 5, at this time 5 is as a motor driven flywheel, transform hydraulic energy into flywheel kinetic energy storage. If the car recovered energy is greater than the set maximum kinetic energy of the flywheel, then the clutch is disengaged. At this time 8 solenoid valve is energized, the excess energy is stored in the accumulator, if the system pressure exceeds the upper limit, the remaining energy release through the relief valve. When the vehicle accelerates, the 3 pump/motor as a motor, high pressure oil power the wheels, the system pressure is reduced. 5 Pump/motor as a pump, the system pressure is kept to a level, high pressure oil flow through the outlet 5 to the bottom of the entrance 3. When speed of the flywheel is in minimum speed allowed, the 6 clutch and connecting the total driveline 11 clutch disconnect, ECPS system does not provide power. 9 Accumulator prevent the system pressure too volatile and maintain a substantially constant system pressure.

3 ECPS regenerative braking principle test system

Verify ECPS regenerative braking system energy recovery efficiency, the vehicle system can provide maximum braking torque is 1/16 to 1/8 for the calculation of the test standard. ECPS equivalence principle test device structure shown in Figure 2.
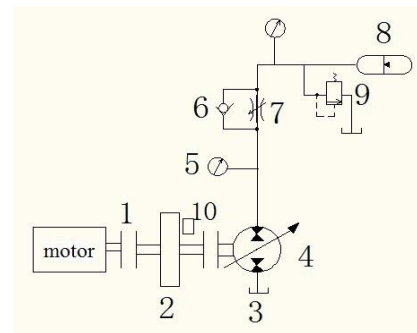
1. Electromagnetic clutch 2. Flywheel 3.Tank 4.Variable pump / motor 5. Manometer 6. Check valve 7. Throttle 8. Accumulator 9. Relief valve 10. Wheel speed sensor

FIGURE 2 Test apparatus schematic diagram of hydraulic regenerative braking system

Its equivalence principle is: When the motor drives the flywheel, the clutch between flywheel and motor disconnected, this time representing the flywheel as energy components provide energy for the hydraulic system. Pump/motor as a pump regenerative braking energy, transforms mechanical energy into hydraulic energy can be stored in the accumulator until the flywheel stops; the accumulator release energy after its pressure stabilized, the pump/motor as a motor transforms hydraulic energy into mechanical energy stored in the flywheel. Flywheel as the energy storage element in the actual pure electric vehicles, along with the hydraulic system and electric power systems arranged in parallel chassis. The corresponding processes are when vehicle accelerating driving conditions and driving conditions when braking. Displacement of pump/motor as experimental variables, measured final pressure of accumulator and final speed of the flywheel, calculated the efficiency of energy storage about accumulator, flywheel energy storage efficiency and hydraulic energy efficiency of regenerative braking system, as the test results shown in Table 1.

TABLE 1 Braking energy recovery test results

	Displacement of pump/motor (L)	Initial kinetic energy of vehicle (J)	Final shaft speed of flywheel (r/min)	Energy store in flywheel (J)	Energy utilization of ECPS system
1	85	167580	1200	39844	23.8%
2	75	167580	1410	55134	32.9%
3	95	167580	1550	66529	39.7%
4	105	167580	1795	89320	53.3%
5	115	167580	1910	101050	60.3%
6	125	167580	2130	125685	75.0%
7	135	167580	1790	88817	53.0%
8	145	167580	1460	78763	47.0%

3.1 EXPERIMENTAL INVESTIGATION

When displacement of hydraulic pump/motor 5L, according to Figure 3, connect 1clutch between the flywheel and motor, then disconnect clutch between 4

529

pump/motor and flywheel, close throttle, after the rotational speed of the motor driving the flywheel to 1500r/min, speed keep stability for 1 min. Then connect the pump/motor and flywheel clutch, disconnect 1 and 4 as a pump, high pressure system oil is stored in the accumulator. When flywheel stop, the accumulator's pressure is 3.61Mpa. Energy stored in accumulator is E1=3638J, the initial kinetic energy of system is E0=7395J, recovery efficiency of the accumulator is $\eta 1$ = 49.2%.
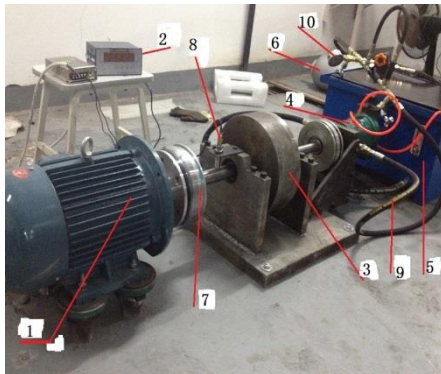
1. Motor 2. Tachometer 3. Flywheel 4.Variable pump/motor 5. Tanks 6. Accumulator 7. Electromagnetic clutch 8. Wheel speed sensor 9. Pipes 10. Manometer
FIGURE 3 ECPS test device

Changing the pump/motor displacement, do in the same test method, the measured results of the other seven sets of data shown in Table 1.

After the accumulator pressure is 3.61Mpa stable, open the throttle, connect the clutch between flywheel with 4 pump/motor, disconnect 1, using the stored energy in accumulator drives flywheel, 4 working as a motor at this time , wheel speed sensor records the flywheel in the process to achieve the maximum speed of 755r/min. The kinetic energy of the flywheel is $E2$=1784J. Flywheel energy recovery efficiency is of

51.5%.

In this same test method, the measured results of the other seven sets of data shown in Table 1.

The ratio of final kinetic energy of the flywheel to initial kinetic energy equals the hydraulic regenerative braking energy efficiency of the total system [6].

$$\eta = \frac{E_2}{E_0} = 25.3\%$$

In the same test method, the measured results of the other six groups of data shown in Table 1.

3.2 ANALYSIS OF TEST RESULTS

Experiments show that: with different displacement of hydraulic pump/motor, the final recovery of the energy and recovery efficiency of flywheel and accumulator is different, the greater displacement is the more energy is recovered. But when pump/motor displacement above a certain value, because the pump/motor resistance increases, energy recovery efficiency of hydraulic brake energy recovery system will decline; Further data indicate that accumulator energy recovery efficiency is lower than recovery efficiency of flywheel. Hydraulic system requires a good match between the components to achieve optimal energy recovery results.

4 System simulation analysis

Based on Figure 1 shown a hydraulic regenerative braking energy recovery system, created simulation model of the desired vehicle system on AMESim [7]. Vehicle simulation model shown in Figure 4, in which the hydraulic energy recovery system shown in Figure 5. Driver's control system parameters are divided into acceleration and braking control parameters, the control type is PID control [8].
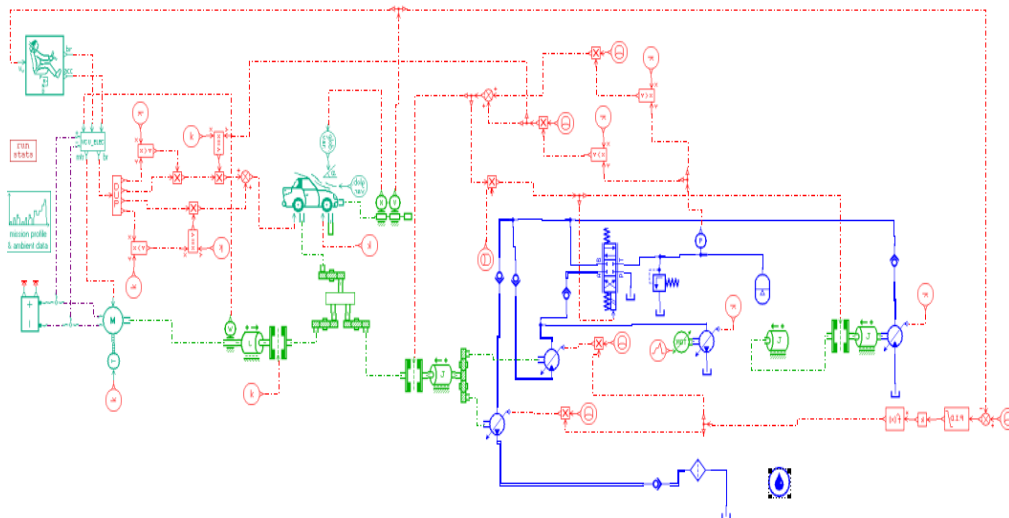

FIGURE 4 The simulation model of vehicle powertrain system with hydraulic regenerative braking system
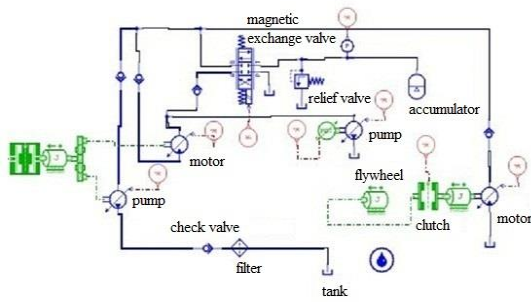
530

FIGURE 5 ECPS hydraulic system model

One model of electric vehicles as the research platform and determine the hydraulic regenerative device system parameters. Assuming that the vehicle goes straight, pump/motor drives alone, calculate displacement of pump/motor results. Make sure the flywheel size parameters: flywheel radius is 250 mm, flywheel thickness is 30 mm, maximum rotate speed is 2500 r/min. Since the hydraulic accumulator system pressure is substantially constant, the fluctuations between 20-25MPa. Select accumulator about the text minimal inflation pressure is 18MPa [9]. Calculate the total volume of the accumulator to be 10L.

4.1 COMPOSITE REGENERATIVE BRAKING SYSTEM BRAKING CONDITION

According to Figure 4 shows a simulation model and vehicle simulation parameters be calculated, then do the simulation of vehicle braking condition. Fully loaded vehicle driving conditions: initial pressure accumulator be set 18Mpa, let hybrid electric vehicle accelerate to 50km/h in 20s, when the time is 30s begins to brake and park. Simulation purpose is to verify the ECPS system and its feasibility of the control system; changing pump/motor displacement to calculate braking energy recovery efficiency under different displacement of pump/motor, and comparison with the corresponding test results. When the brake intensity is $0 < z \leq 0.3$, simulation results shown in Figure 6.

When the brake intensity is $0 < z \leq 0.3$, Figure 6a shows vehicle can decelerate smoothly to stop when braking , driving state and control target is close, means good control effect. The solid red line in Figure 6b represents the secondary component during acceleration and constant speed don't work in the 0-30s, accelerating only by a motor torque which is represented in green solid line; in 30-45s secondary component recovered braking energy when braking, suppose rear tires without friction brake, braking force provided by the ESPS system only. Blue curve presents the sum of the rear axle torque, means motor torque in 0-30s and the toque after secondary component torque multiplied by the gear ratio in the 30-45s.
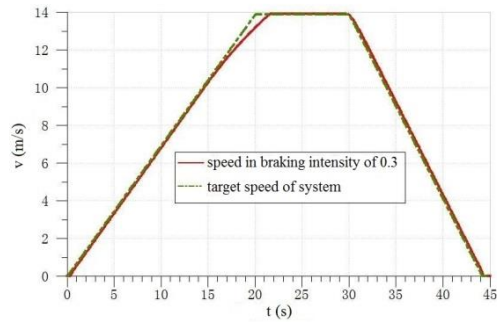
a) Vehicle speed comparison chart
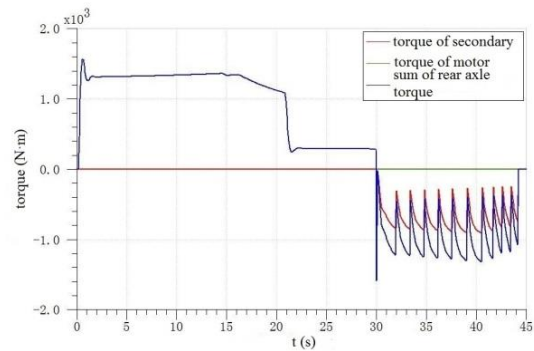
b) The rear axle torque
FIGURE 6 The vehicle braking condition when $0 < z \leq 0.3$

Figure 7 is the change of flywheel rotate speed when change displacement of the pump/motor between 75L-145L.
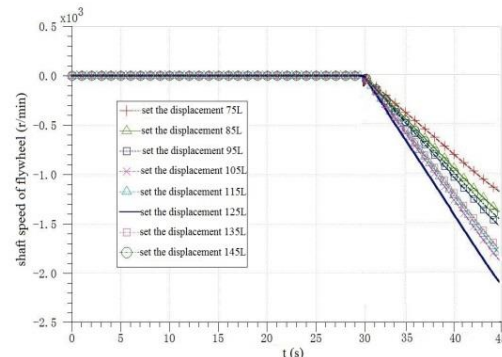
FIGURE 7 The flywheel shaft speed change

Figure 7 shows the rotate speed of flywheel can be affected by pump/motor displacement size. When the pump/motor displacement is small, flywheel rotate speed is increased with the increasing of displacement, and the more stored energy; but when the pump/motor displacement is above a certain value, flywheel rotate speed will reduce with increasing of pump/motor displacement, energy recovery is also reduced. From changing rotate speed of the flywheel can know how much energy storage in flywheel, according to the initial kinetic energy of vehicle can obtain energy recovery efficiency of the hydraulic regenerative braking system, the results shown in Table 2.

531

TABLE 2 Energy recovery efficiency of hydraulic regenerative braking system

	Pump/motor displacement (L)	Initial Shaft speed of flywheel (r/min)	Final speed of flywheel (r/min)	Final pressure of accumulator (MPa)	The efficiency of energy storage of accumulator	The efficiency of energy storage of flywheel	Energy utilization of ECPS system
1	5	1500	755	3.61	49.2%	51.5%	25.3%
2	7	1500	775	3.69	50.2%	53.4%	26.8%
3	9	1500	781	3.74	51.2%	53.0%	27.1%
4	10	1500	812	3.80	52.3%	56.0%	29.3%
5	11	1500	820	3.82	52.7%	56.7%	29.9%
6	12	1500	809	3.79	52.1%	54.7%	28.5%
7	14	1500	780	3.75	50.9%	51.9%	26.4%
8	15	1500	774	3.70	50.7%	50.7%	25.4%

Figure 6a shows, the simulation condition is consisted by acceleration and braking conditions, corresponding to the release energy and regenerative braking energy of ECPS system. Since the test parameters for the vehicle simulation parameters obtained according to certain scaling, so during the simulation and testing process will let pump/motor displacement as variables, respectively get the impact of pump/motor displacement for energy efficiency of regenerative braking system (shown in Table 2 and Table 1 ECPS system energy efficiency), the test results and simulation results can be verified with each other. In the test, the pump/motor variable range for 5-15L; in the simulation study, the pump/motor variable range for 75-145L.

By comparing Tables 1 and 2, when pump/motor displacement is small, the flywheel speed is increased with the increase of displacement and the more stored energy. But when the pump/motor displacement value higher than a certain time (in Table 1, this value is about 11L; simulation shown in Table 2, this value is about 125L), the flywheel speed will decreases with the increases of pump/motor displacement, also the recovered energy is reduced. Mainly due to resistance increases of the pump/motor, after displacement higher than a certain value, the energy recovery efficiency will decline, that indicate secondary component parameters need to good match to get the best effect of energy recovery.

Although tests can verify the system can recover some energy during braking (ECPS system as shown in Table 1, the highest energy efficiency is about 30%), but this is difference between energy recovery efficiency and simulation result (as Table 2 shows ECPS system energy use rate of 75%). Mainly reason in the test:

1) air resistance and bearing resistance caused energy loss during flywheel rotating in high speed, the bearing resistance is a larger proportion of energy loss account when flywheel in low speed;

2) there may be leaks in high pressure pipes, throttle adjustment is manual adjustment, there are some errors;

3) hydraulic components need to optimize matching.

4.2 COMPOUND REGENERATIVE BRAKING SYSTEM DRIVING CYCLE SIMULATION

To test the effect of ECPS energy recovery system, this paper according to GB/T 18386-2005 specified urban driving cycle (condition ECE-15) to work conditions simulation [10], worked for 800s simulation in urban conditions. Simulation results are shown below.
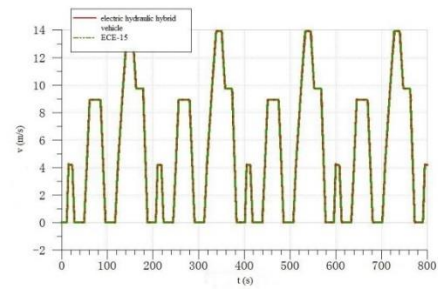
a) Driving- cycle speed comparison
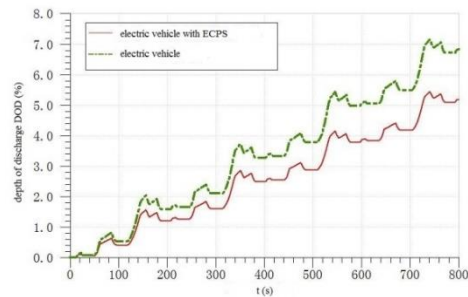
b) Depth of discharge comparison
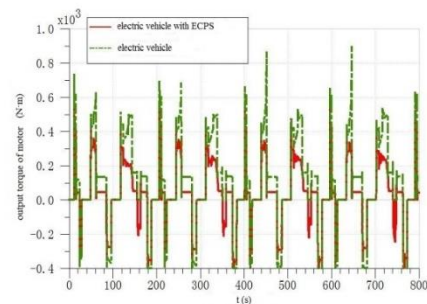
c) Output torque of motor comparison
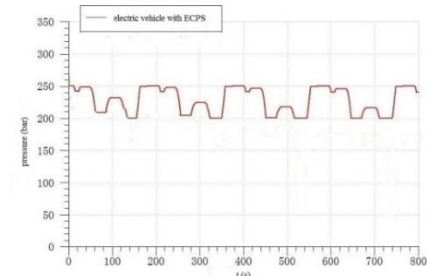
d) System pressure comparison.
FIGURE 8 ECE-15 driving cycle simulation results

Figure 8a shows that vehicle speed curve of electro-hydraulic hybrid is basically close to speed curve of ECE-15 conditions, it's means that the braking efficiency of ECPS system is stability; Figure 8b shows, the depth of discharge of ECPS electro-hydraulic hybrid vehicle lower than pure electric vehicles (ECPS electro-hydraulic hybrid vehicles is 5.11%, pure electric vehicles 6.88%), the battery life can be increased accordingly. In terms of percentage calculations, ECPS electro-hydraulic hybrid vehicles compare to pure electric vehicles, driving range can increase about 25 %.

Figure 8c shows, motor load and load change of ECPS electro-hydraulic hybrid vehicle is small, basically the motor is working at rated torque , with little need to peak torque output; Figure 8d shows, ECPS system pressure is stable can avoid high pressure impact hydraulic components and good system dynamic performance.

5 Conclusions

Studies show that electro-hydraulic hybrid vehicle with ECPS's driving range increased by 25 % compared with electric vehicle without ECPS; Because of motor load decreases, reduce the depth of battery discharge, so battery life can be extended. The main factors impact recovery efficiency of ECPS braking energy is the pump/motor displacement, and how to control the pump/motor displacement change in a variety of braking intensity to achieve energy efficient recovery, there is reference value about the results of this study.

Acknowledgments

References

[1] Eriston L L, Miles M D 2008 Retrofittable regenerative braking in heavy application *SAE* (08) 33-42
[2] Bulter K L, Kamath P A 1999 *IEEE Transactions on Vehicular Technology* **48**(6) 1770-8
[3] Peng D, Zhang Y, Yin C L 2007 Design of hybrid electric vehicle braking control system with target wheel slip ratio control *SAE* **01**(1515)
[4] Yasushi A, Takaomi, Shirase 2007 Development of hydraulic servo brake system for cooperative control with regenerative brake *SAE* **01**(0868)
[5] Nakazawa N 1987 Development of a braking energy regenerative system for city buses *SAE* (872265)
[6] Tao N, Jihai J, Hui S 2009 Hydraulic hybrid vehicle regenerative braking Strategy based on backward modeling parallel *China Mechanical Engineering* **20**(15) 1880-4 (*in Chinese*)
[7] XU Y, Ning X, Wang Q 2012 Simulation analysis of hydraulic regenerative braking system for pure electric vehicle based on AMESim *Mechanical and Electrical Engineering* **29**(2) 1068-72
[8] Qu Y, Shu H, Xiong S 2010 PID neural network improvement *Electrical and Mechanical Engineering Technology* **39**(8) 39-41
[9] Zhang Z 1997 Hydraulic drive system *Beijing Mechanical Industry Press*.
[10] Zhao J 2009 Research on Strategy of Regenerative Braking and Coordinated Control for Hydraulic Energy-saving Vehicle *Jilin University*

Authors

Xiaobin Ning, born on July 13, 1965, Shanxi, China

Current position, grades: associate professor.
University studies: University of Science & Technology Beijing.
Scientific interest: vehicle dynamic simulation and control.
Publications: Prediction for Shoe Factor of Drum Brakes Based on Nonlinear 3D Simulation Models. Applied Mechanics and Materials, 2010, 38: 880-5.
Co-simulation of Wheel Loader Working Mechanism. Applied Mechanics and Materials, 2010, 43:72-7.
Design and Simulation of the Suspension System of Chassis Platform Based on Handling Stability. SAE Paper No.2010-10-0723.
Simulation Based Design for Heavy Truck Brake. SAE Paper No.2009-01-0581.
Co-simulation of Steering Mechanism of Truck. SAE Paper No.2008-01-1104.

Ning Li, born on February 9, 1986, Shanxi, China

Current position, grades: junior lab technician.
University studies: Zhejiang University of Technology.
Scientific interest: automotive dynamics.
Publications: 4 papers.

Junping Jiang, born on January 6, 1983, Shaanxi province, Weinan city, China

Current position, grades: intermediate automotive engineer.
University studies: Zhejiang University of Technology.
Scientific interest: automotive dynamics.

Thrust line measurement of multi-nozzle solid rocket engine

Chunfu Zhang[1]*, Jianguo Miao[2], Song Tang[2]

[1]*School of Electrical Engineering, YanCheng Institute of Technology, Yancheng 224051, Jiangsu, China*

[2]*Chenguang Calibration & Testing Center, Nanjing 210006, Jiangsu, China*

Abstract

Static thrust line is an important measuring parameter of solid rocket engine. Factors that affect the misalignment and deflection of thrust line of multi-nozzle solid rocket engine were analysed in static condition, and a measuring method by Laser Tracker for thrust line was researched. A calculation model of the thrust line was established based on simplified force spiral rule, which synthesizes the space force systems. The simulation results of misalignment and deflection indicate that the method has high measurement accuracy.

Keywords: solid rocket motor, static thrust line, multi-nozzle, laser tracker

1 Introduction

Before shipment, various geometric parameters of solid rocket motor need to be measured accurately under static conditions. Misalignment and deflection of the engine's thrust line are very critical parameters. The misalignment indicates the vertical distance between the thrust line and the engine axis in space. The deflection refers to the angle of the vector with the engine axis of the thrust line in space, the two parameters directly affect the pitch and yaw of the rocket during the flight to control the attitude [1].

Thrust of rocket engine is generated like this, jet flow from burning huge quantities of fuel acts upon internal curved surface of the nozzle, therefore static thrust line can be substituted by the geometry centreline of the surface. The thrust line of multi-nozzle solid rocket engine can be obtained by synthesizing the thrust produced by each nozzle, while the factors affect the thrust line are due to the deviation relative to the theoretical position of nozzles in the engine. During the machining and assembling process, the deviations of engine nozzle parts exists inevitably, and owning to the accumulation and amplification, will affect the action points and direction of each thrust line. Thus, the misalignment and deflection will be brought in the synthesis of thrust line eventually [3, 4].

In the past for a long time, under static conditions, the thrust line of a single nozzle is substituted by the connection from the centre of the throat neck and the centre of the nozzle outlet end face. Meanwhile, in multi-nozzle engine, the synthesis of the thrust line is calculated based on the arithmetic average of each centre of the throat neck and the nozzle's outlet end face indirectly, apparently there will be deviation in principle. In [5], we have already researched the measuring technology of a solid rocket engine with single nozzle's structure. In this letter, a four-nozzle solid rocket motor is taken as a measuring object, the thrust line measurement on multi-nozzle engine and

data processing methods are studied further to achieve efficient and accurate results.

2 Static thrust line measurement of single nozzle rocket engine

The measurement of a single nozzle's thrust line can be realized by surveying the geometric centreline of the internal surface, which is determined by the centres of the nozzle throat neck and outlet. In the geometric measuring point of view, typically, between the nozzle throat neck and outlet end face, a number of cross-sections of the nozzle parallel to the outlet plane are needed to be measured in order to improve accuracy, and then the geometry centreline is fitted from these sections' centres by least square method. The measurement diagram is shown in Figure 1.

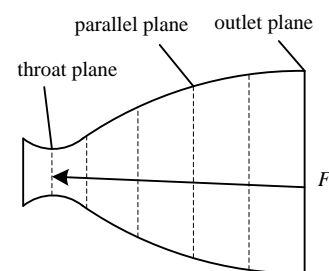FIGURE 1 Diagram of thrust line measurement

After fitting, the space equation of the thrust line *F* can be expressed as Equation (1) shows.

$$\frac{x - x_1}{l_1} = \frac{y - y_1}{m_1} = \frac{z - z_1}{n_1} \ . \tag{1}$$

For multi-nozzle solid rocket motor, the same method can be applied to get the other nozzles' thrust line equations.

Corresponding author e-mail: zhangchunfu@ycit.edu.cn

3 Static thrust line analysis of multi-nozzle rocket engine

According to the force translational theorem of rigid body in theoretical mechanics, any force system acting on the same rigid body can be simplified to any point and synthesized into a resultant force and an additional spatial force couple. In multi-nozzle engine, the action points and direction of each sub-thrust are different; in order to calculate the misalignment and deflection of the synthesis engine thrust line, each nozzle's thrust vector needs to be simplified and synthesized of the space force.

Assume that the direction number of the thrust line F_1 is (l_1, m_1, n_1), the action point's coordinate is (x_1, y_1, z_1); the direction number of the thrust line F_2 is (l_2, m_2, n_2), the action point's coordinate is (x_2, y_2, z_2); the direction number of the thrust line F_3 is (l_3, m_3, n_3), the action point's coordinate is (x_3, y_3, z_3); the direction number of the thrust line F_4 is (l_4, m_4, n_4), and the action point's coordinate is (x_4, y_4, z_4), the distribution of each thrust's vector is shown in Figure 2, the illustrated coordinate system is the product coordinate system of the engine.
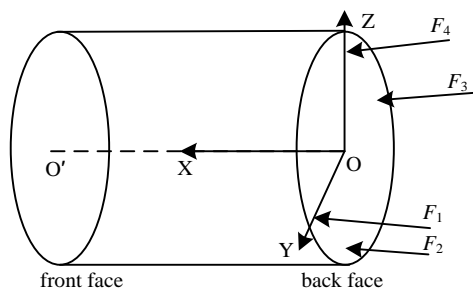
FIGURE 2 Diagram of thrust forces distribution of multi-nozzle

Decompose the thrust component along each coordinate axis of the product coordinate system and then synthesize them in each axis direction separately, the component of the resultant force can be obtained as shown in Equation (2).

According to the calculation by Equation (2), all the force components along the coordinate axis direction can synthesize the magnitude and direction of the resultant thrust line Fc based on projection and analysis.

$$\begin{cases} F_x = \sum_{i=1}^{4} F_i l_i \\ F_y = \sum_{i=1}^{4} F_i m_i \\ F_z = \sum_{i=1}^{4} F_i n_i \end{cases} \quad (2)$$

The magnitude and direction of Fc is shown as below in Equation (3).

$$\begin{cases} F_c = \sqrt{F_x^2 + F_y^2 + F_z^2} \\ l_{F_c} = \dfrac{F_x}{F_c} \\ m_{F_c} = \dfrac{F_y}{F_c} \\ n_{F_c} = \dfrac{F_z}{F_c} \end{cases} \quad (3)$$

When each component of thrust is simplified to the origin of coordinate system, the additional space moment of couple is shown as below in Equation (4).

$$M_x = \sum_{i=1}^{4} F_i \left(n_i y_i - m_i z_i \right)$$
$$M_y = \sum_{i=1}^{4} F_i \left(l_i z_i - n_i x_i \right) \quad (4)$$
$$M_z = \sum_{i=1}^{4} F_i \left(m_i x_i - l_i y_i \right)$$

According to the calculation by Equation (4), the synthesis results and direction of the resultant additional couple moment Mc also can be calculated based on projection and analysis as Equation (5) shows.

$$\begin{cases} M_c = \sqrt{M_x^2 + M_y^2 + M_z^2} \\ l_{M_c} = \dfrac{M_x}{M_c} \\ m_{M_c} = \dfrac{M_y}{M_c} \\ n_{M_c} = \dfrac{M_z}{M_c} \end{cases} \quad (5)$$

Due to the influence of the deviation in machining and assembling process, each nozzle thrust of multi-nozzle solid rocket motor in space inevitably acts as non-converging system of forces, and cannot be a parallel forces system. Thus, the results of the thrust component simplified to the origin of product coordinate system need to be further simplified to synthetic force screw, as shown in Figure 3, then the principal vector Fc is the physically meaningful synthetic thrust line.

The synthesis process of force screw is as follows: first build the auxiliary space coordinate system parallel to Fc according to Equation (3), project and decompose the additional moment of couple under the coordinate system built above on the basis of Equation (5), the relative coordinates of the main vector which need to be translated can be obtained in the force spiral synthesis process. Then in accordance with the spatial relationships between the products and auxiliary coordinate systems, convert translational coordinates of the principal vector into the product coordinate system, the coordinates of the feature points of the principal vector equation can be acquired, and finally the equation of synthesis thrust line is achieved.

The process is realized in specific data-processing software based on Visual Basic language.

FIGURE 3 Diagram of simplified force screw

Based on the theory above, the line distance equation and angle equation in space, the misalignment d and the deflection θ of the synthetic thrust line can be calculated in Equations (6) and (7).

$$d = \frac{\begin{vmatrix} x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix}}{\sqrt{\begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix}^2 + \begin{vmatrix} n_1 & l_1 \\ n_2 & l_2 \end{vmatrix}^2 + \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix}^2}}, \tag{6}$$
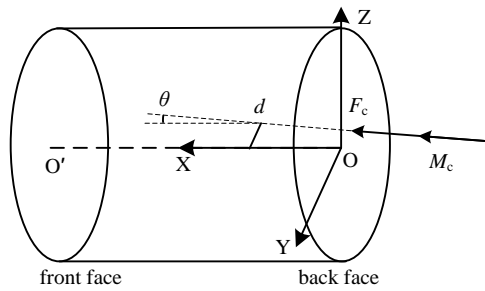
$$\theta = \arccos\left(\frac{l_1 \cdot l_2 + m_1 \cdot m_2 + n_1 \cdot n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2} \cdot \sqrt{l_2^2 + m_2^2 + n_2^2}}\right). \tag{7}$$

4 Simulation

Firstly, the engine product coordinate system is defined, the X-axis is the centreline from the engine front and rear skirts of the end face, i.e. the engine reference axis, the Y-axis is determined by the first quadrant of engraved line of the rear skirt, and the Z-axis is determined by the right-hand rule.

Secondly, according to the definition of the thrust line, use the internal grate of the nozzle as the reference plane, measure 8 uniform sampling points which are set on a circle of the internal throttle of the nozzle in the engine product coordinate system.

Thirdly，according to the definition of the thrust line, use the internal end face of the nozzle outlet as the reference plane, measure 8 uniformly distributed sampling points, which are set on a circle of the outlet end face of the nozzle in the engine product coordinate system [6].

Finally, fit the simulation coordinates obtained above separately, which are on the laryngeal neck and the nozzle outlet end, the fitted centre coordinates is as shown in Table 1.

TABLE 1 Fitting results of nozzle throat and outlet

No.	x (mm)	y (mm)	z (mm)
1	0.12	-365.16	0.21
2	0.04	364.95	0.13
3	-0.09	-0.17	365.16
4	-0.41	0.02	-365.01
5	-400.03	-365.04	-0.13
6	-400.01	365.14	-0.07
7	-399.94	0.11	364.88
8	-399.91	0.05	-365.09

By fitting the results in Table 1 respectively, each space expression of the thrust component can be obtained.

In accordance with the force screw synthetic rule, the space equation of the synthetic thrust line (to ensure the accuracy of calculation, the coordinates of the feature points are retained to six decimal) is as shown in Equation (8).

$$\frac{x}{1} = \frac{y + 0.112834}{-0.000388} = \frac{z - 0.000114}{0.000563}. \tag{8}$$

Based on Equations (6) and (7), the misalignment and deflection of the thrust line are separately calculated, the results is illustrated in Table 2. Other processing results are listed in the table simultaneously, which use geometric mean method based on centre points of the end face of the throttle and nozzle outlet.

TABLE 2 Results of thrust line measurement

Method	misalignment (mm)	deflection (′)
force screw	0.093	2.349
geometric mean	0.005	2.349

From Table 2, we can see that the two methods have a large deviation between the misalignments, while no deviation between the deflections of the thrust line. This is because the decomposition and synthesis operation to the space force system axis has a linear relationship with the geometric mean arithmetic, and has the same effect in space; while in the geometric mean method, the role of space additional couple is not considered, thus a larger schematic deviation exists.

5 Conclusions

On the basis of the measurement technology of a single nozzle solid rocket motor's thrust line, multi-nozzle static thrust line measurement techniques were researched using laser tracker, a space geometry model of synthetic static thrust line was established according to the synthetic force screw rule, and the equation was given. The simulation showed that the measuring techniques we introduced could overcome the principle error of the geometric mean method, and had higher accuracy.

References

[1] Zhu C, Xu X 2009 The Dynamic Measurement of Thrust Misalignment in Solid Rocket Motors Based on Load Identification Technique *The Ninth International Conference on Electronic Measurement & Instruments* Beijing China 544-7

[2] Gorinevsky D, Samar S, Bain J, Aaseng G 2005 Integrated Diagnostics of Rocket Flight Control *IEEE Aerospace Conference* MT USA 1-12

[3] Lai P, Tian W, Yu Z 1999 Simulation Calculation of the Misalignment and Deflection of Thrust Line of Solid Rocket Motor *Journal of Solid Rocket Technology* **22**(1) 35-8 *(in Chinese)*

[4] Taur D R, Chem J S 1997 Optimal Thrust Vector Control of Tactical Missiles AIAA-97-3475 *(in Chinese)*

[5] Zhang C, Tang W, Li H, Wang J, Chen J 2007 Application of laser tracker to thrust line measurement of solid rocket motor *Journal of Solid Rocket Technology* **30**(6) 548-51 *(in Chinese)*

[6] Bao F, Guo D, Zhao F, Liu J 2004 The integrated design and analysis of solid rocket motor nozzles. *Journal of Solid Rocket Technology*. **27**(3), 169-172,183 *(in Chinese)*

Authors

Chunfu Zhang, born in October, 1977, China

Current position, grades: lecturer at Yancheng Institute of Technology, China.
University studies: PhD degree in instrument science and technology from Harbin Institute of Technology, China in 2007.
Scientific interests: automated testing and geometrical measurement.

Jianguo Miao, born in September, 1962, China

Current position, grades: senior engineer at Chenguang Calibration & Testing Center, China.
University studies: Bachelor's degree in instrument science and technology from Harbin Institute of Technology, China in 1983.
Scientific interests: mechanical testing and geometrical measurement.

Song Tang, born on May 5, 1986, China

Current position, grades: assistant engineer at Yancheng Institute of Technology, China.
University studies: Master's degree in instrument science and technology from Harbin Institute of Technology, China in 2011.
Scientific interests: mechanical testing and geometrical measurement.

Vibration characteristic researching of capillary of ultrasonic wire bonding

Wuwei Feng*

Department of Ship and Ocean Engineering, Zhejiang Ocean University, Zhoushan, 316022, China

Abstract

Wire bonding, a process of the connection between a semiconductor chip and a lead frame by a thin metal wire, is one of the important processes of electronic packaging. Vibration characteristic and friction behaviour of capillary of Microelectronic Packaging ultrasonic bonding system are studied. A dynamic contact model of capillary was built by finite element method to gain bonding mechanism of wire bonding system. Vibration response and contact friction property of the capillary are calculated. The effects of the loading frequency and static compressive force to the vibration response and contact friction stress were obtained, and the relationships between the loading frequency, the static compressive force and the vibration response were given. The researching conclusions can be used as references for having a better understanding of the bonding mechanism and the study of fault diagnose techniques for bonding process.

Keywords: ultrasonic wire bonding, contact analysis, finite element, capillary

1 Introduction

Wire bonding is a process for connection between a semiconductor chip and a lead frame by a thin metal wire [1-3]. For every error bonding point will cause the IC circuit failure entirely, thorough researches about bonding mechanism and the effect factor of bonding quality are indispensable. Capillary is a key device of bonder, so the research on its dynamic characteristic and contact friction behaviour will help to master bonding mechanism well. For capillary's high-speed and non-linear vibrate characteristic, it is very difficult to study bonding mechanism through routine methods. In this paper finite element method is presented to investigate the vibration characteristic of capillary. First, the mode of capillary is analysed and first nature frequency is got, second a dynamic contact model is built and referring to the parameter, got from mode analysis, the load to capillary is confirmed. The effect factor of contact friction stress is acquired in bonding process. The result can be used as references for having a better understanding of the bonding mechanism and the study of fault diagnose techniques for bonding process [4-8].

2 Ultrasonic wire bonding system introduction

Ultrasonic wire bonder (shown as Figure 1) by and large consists of a phase-locked-loop (PLL) ultrasonic generator, an ultrasonic transducer system (a piezoelectric driver, a barrel, a concentrator and a capillary), a work holder to support the materials to be bonded, and means for applying a static compressive force to the materials being bonded [1]. The ultrasonic, generated from PPL ultrasonic signal generator, is transformed into high frequency mechanical vibration energy through piezoelectric transducer. Over the years, various models have been postulated in describing the possible mechanism involved in ultrasonic wire bonding [7, 9].

For quality control purposes, it is convenient to model the mechanism as follows: the materials being bonded are first brought into contact under an external static compressive force, producing some initial deformation but no adhesion due to the presence of surface films or oxides. Energy is then introduced via the transducer, which vibrates at ultrasonic frequency, thus it results in oscillatory forces parallel to the wire-pad interface. The wire softens upon absorption of energy and flows under loading, breaking up the surface oxides and exposing a fresh surface of both the wire and the boding pad. Interfacial movement is progressively restricted by the formation of multiple micro welds in the peripheral area of the contact zone. Continued application of cyclic stresses produces plastic deformation of wire [1].

Figure 1 shows that the system is a strong coupling connection with tribology and dynamics. Bonding process is a very short time, especially for high-speed bonder. Capillary is a key device of wire bonder, so in this paper its dynamics characteristic and bonding mechanism are studied through finite element method.
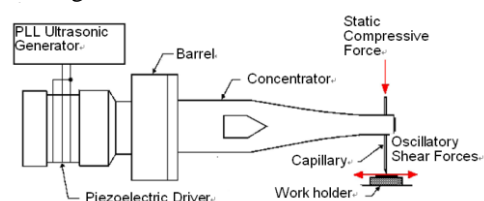

FIGURE 1 Structure of an ultrasonic wire-bonding system

---
* *Corresponding author* e-mail: fengwuwei@163.com

3 Model of capillary

In order to accurately reflect bonding system properties and simultaneously simplify system model, only a model of capillary is built here. Vibration energy, imputed from concentrator, is loaded to the tail end (7.73 mm to 11.1 mm) of capillary through a function as Equation (1) (see Figure 2). A work holder is built under the capillary to study contact friction behaviour between capillary and work holder [2].
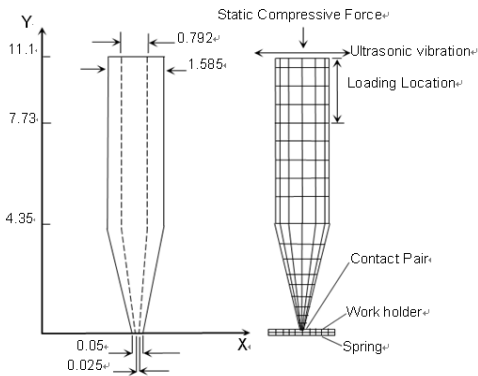
FIGURE 2 Size and model of capillary/mm

Capillary is a high-density and high hardness ceramic material. Solid 45 is selected for capillary and work holder and Combine 14 for spring. Refined mesh is generated at the tip of capillary to accurately reckon its deformation. Target element and contact element are selected Target 169 and Contact 171 respectively [5]. In order to simulate real movement and restraint of capillary, only x direction freedom vibration of capillary's top (7.73 mm to 11.1 mm) is allowed. The equation is:

$$F = 1 \times 10^{-6} \times \sin(2\pi\omega t) , \qquad (1)$$

Frequency range from 60 kHz to 68 kHz, size of static pressure is from 0.02 N to 0.2 N, and computation time is 10ms at a time interval of 2×10^{-3} ms. Beside y direction freedom of work holder, others are restrained.

4 Vibration response of capillary

4.1 FREE VIBRATION ANALYSIS OF CAPILLARY

Capillary is multi-DOF cantilever structure. It is difficult to extract the intrinsic mode through theoretical or experimental methods. In this paper inherent modal are calculated and extracted by the finite element method. Finite element analysis software, ANSYS80, is used to calculate its mode and dynamic response. Here, we are more interested to extract the mode range from 0 to 300 kHz. Table 2 shows the other order mode of capillary. We may see that mode of capillary about x axis and y axis is same as the structure of capillary is symmetry.

Mode deformations of capillary are shown in Figure 3. Figures 3a, 3b, and 3c show the deformations of first order, second order and torsion respectively. The first order nature frequency about x-axis is 65,140 Hz. It is very

adjacent with work frequency of capillary. The mode of only x-axis or y-axis direction is inspired for concentrator's horizontal vibration.
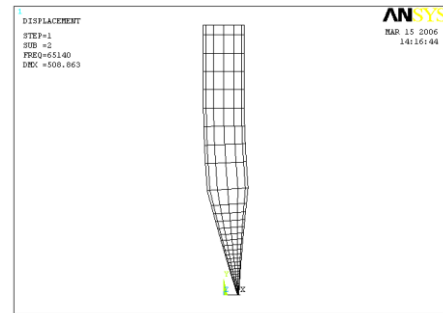
a) First order mode
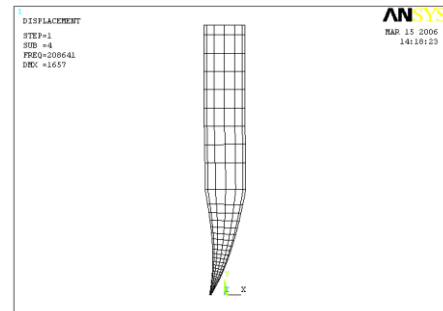
b) Second order mode
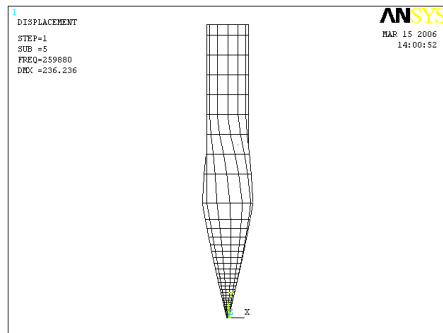
c) Torsion mode
FIGURE 3 Mode deformation of capillary

4.2 SELECTING FRICTION MODEL OF CONTACT ANALYSIS

Bonding process is a flexible contact process, which is still pending further study. So far, there is still no universally agreed specific solution. For a long time, many dynamicity and mathematicians have been working hard to solve the contact problem. With the development of computer science, numerical-based and similar-deal approach are rapidly developed. Finite element method is the main tool among them. Here, a contact pairs must be defined where contact might occur during the deformation of model. Once potential contact surfaces are identified, we can define them as target and contact elements, which will the track the kinematics of the deformation process. Target and contact elements make up a contact pair are associated with each other via a shared real constant set (Figure 4).

In the basic Coulomb friction model, two contacting surfaces can bear shear stresses up to a certain magnitude

across their interface before they start sliding relative to each other. This state is known as sticking. The Coulomb friction model defines an equivalent shear stress τ, at which sliding on the surface begins as a fraction of the contact pressure p ($τ = μp + COHE$, where $μ$ is the friction coefficient and COHE specifies the cohesion sliding resistance). Once the shear stress is exceeded, the two surfaces will slide relative to each other. This state is known as sliding. The sticking/sliding calculations determine when a point transitions from sticking to sliding or vice versa [5].
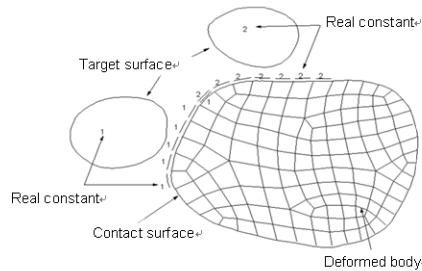
FIGURE 4 Localized contact zones

5 Result of calculation

According to vibration theory, case 1, amplitude response of capillary's tip will be changed at different loading frequencies. Amplitude response of capillary's tip is the biggest under natural frequency. Considering energy point, the system energy inspired will be the biggest when loading frequency is natural frequency. Case 2, when the static pressure is changed, amplitude response of capillary's tip will also be changed. The reason is that bonding system is a damping system and friction change will affect damping coefficient. With damping increasing, the amplitude response will reduce, vice versa. Much static pressure will be carried in order to improve bonding quality, but over high static pressure will also arouse destruction. In this paper, the effect relations between the amplitude response and frequencies or static pressure are studied. The deformation of capillary under loading is shown as Figure 5.
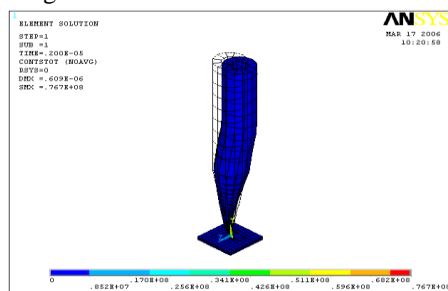
FIGURE 5 Vibration deformation of capillary

5.1 AMPLITUDE RESPONSE WITH DIFFERENT FREQUENCIES

Here, we have more interests about the amplitude response of capillary tip and its friction situation. So some nodes at the capillary tip are selected to study bonding mechanism. The amplitude response of capillary tip with loading frequencies from 63,140 Hz to 67,140 Hz (where static pressure is 0.02 N, frequency interval of 1,000 Hz) is shown in Figure 6. We can see from the figure that the amplitude response increase straight until it become stable under natural frequency, while there are continuous wave when loading frequencies are deviated with natural frequency. Furthermore, with the incensement of deviating, the amplitude response will be gradually reduced.
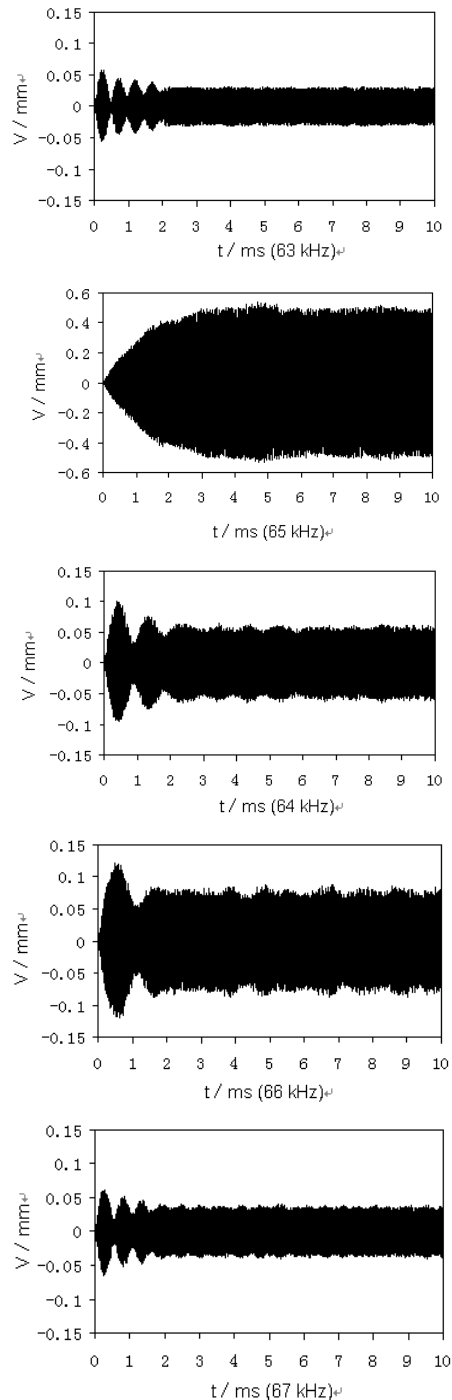
FIGURE 6 Applications response of capillary's tip

The curves between amplitude response and loading frequencies are shown in Figure 7. Figure 7a denotes the change that the amplitude response reach stable, Figure 7b is the biggest of amplitude response with different loading frequencies. As expected, there is a peak-amplitude at the 65 kHz and it is showed in the harmonic response analysis. The stable amplitude is same with the amplitude extremum under the natural frequency, while the earlier is smaller than latter under the others frequencies. Furthermore, the amplitude fluctuant are severe when run at non-natural frequency. It will affect bonding quality seriously.

The contact friction stresses are also different when different loading frequencies are adopted. The curve between maximum contact friction stresses and different loading frequencies (where the same static pressure is used as 0.02 N) is shown in Figure 8. We can see from Figure 8 that the maximum contact stress appears at the natural frequency. When the loading frequencies are increased or reduced, contact stresses are smaller than that under natural frequency. Bonding process is carried out to use friction stress mainly, so it is expected that the work frequency should be most close to the natural frequency.

5.2 AMPLITUDE RESPONSES WITH DIFFERENT PRESSURES

Contact stress and amplitude response of capillary tip are changed under the static pressure is different. Insufficient static pressure will result in shortage of contact stress, at the same time, exorbitant one cannot be achieved fairly good results. Figure 9 shows the curves between loading pressure and amplitude response of capillary tip (where there are five curves, they denote five different loading frequencies, respectively). Amplitude responses of capillary tip are reduced with bonding force increasing. The minimum amplitude of natural frequency is bigger than the maximum amplitude of other frequencies.
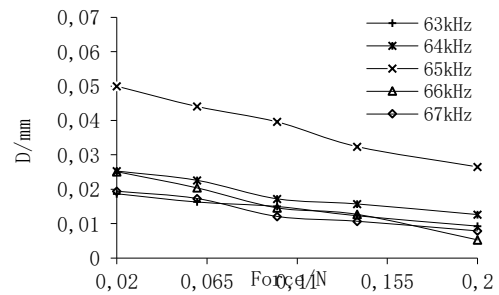
FIGURE 9 Applications response versus bonding force (where five curves denote five different frequencies, respectively)

However, the amplitude response will be quite unstable under exorbitant bonding force. Figure 10 shows amplitude response of capillary tip versus loading times under exorbitant bonding force of 2 N. We may see that amplitude response of capillary tip is mostly near to zero sometime. As expected, the capillary tip cannot move due to too high friction.
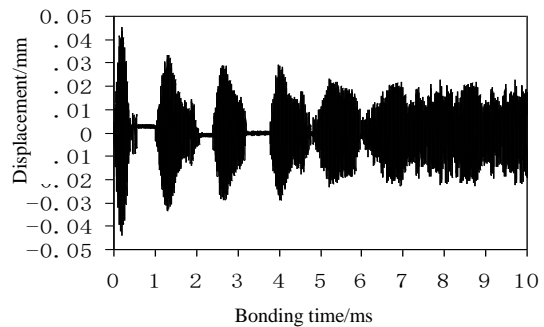
FIGURE 10 Amplitude response versus bonding time (bonding force=2N).

The contact friction stress curve and its spectrum when bonding force is 2 N are shown in Figure 11. Contact stress is increased or reduced periodically. Its trend is very close to Figure 10. There is a peak-amplitude at 120 kHz from frequency spectrum curve. The main reason is that there are two contacts occurring within every vibration cycle.

a)

b)

FIGURE 7 Applications response versus frequencies

FIGURE 8 Maximum contact stress versus frequencies (bonding force=0.2 N)

541
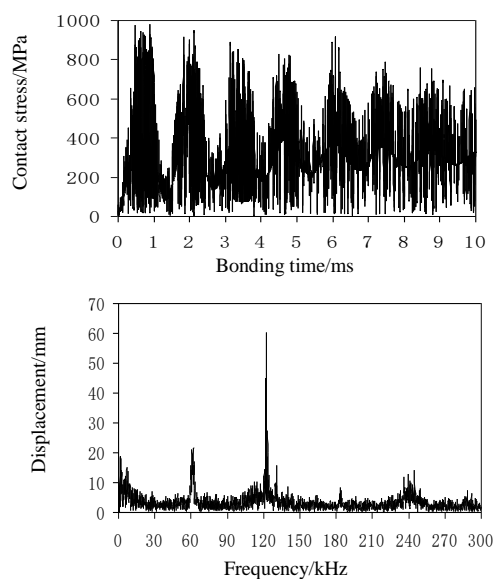
FIGURE 11 Contact stress and frequency spectrum (bonding force=2N).

method to master wire bonding system bonding mechanism. Vibration response and contact friction property of the capillary are calculated. The effects of the loading frequency and static compressive force to the vibration response and contact friction stress were obtained. From above researching works, the following conclusions may be gained: First, amplitude response of capillary relatives with the loading frequencies. Under non-natural loading frequency, amplitude response wave severely. Second, a suitable bonding force must be selected in bonding process. Excess high or low bonding force will affect bonding quality seriously. Third, times to reach stabilization of amplitude response are different under various bonding force. The bigger bonding force, the shorter bonding stable times, vice versa.

6 Conclusions

In this paper, the dynamic characteristics and contact friction behaviour of capillary are studied. A dynamic contact model of capillary was built by finite element

Acknowledgments

References
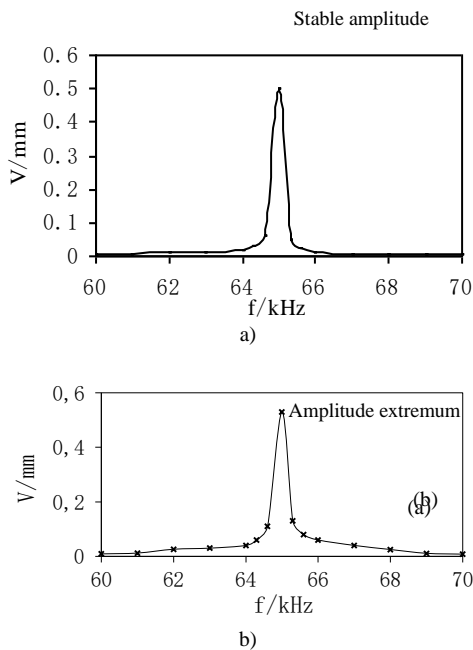
[1] Pufall R 1993 Automatic process control of wire bonding *Proceedings of the Electronics Components Technology Conference* 159-62
[2] Harman 1998 Wire Bonding in Microelectronics: Materials and Processes, Reliability and Yield *McGraw-Hill 2nd Edition*
[3] Hu C M, Guo N Q, Ling S F, Dyu J 1998 *IEEE/CPMT* 202-5
[4] Mao C, Min S 1995 Basic Study of Finite Element Numerical Method *Beijing: Tinghua University Press (in Chinese)*
[5] Or S W, Chan H L W, Lo V C,Yuen C W 1998 *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Contro*l **45**(6) 1453-60
[6] ANSYS Training Manuals 1999 *ANSYS China Acting Company*
[7] Pan Z 2004 Finite Element Analysis and Application *Tsinghua University Press (in Chinese)*
[8] Or S W, Chan H L W, Lo V C, Yuen C W 1996 Sensors for automatic process control of wire bonding *Proceedings of the 10th IEEE International Symposium on Applications of Ferroelectrics (ISAF) East Brunswick NJ USA* **2** 991-4
[9] Or S W, Chan H L W, Lo V C, Yuen C W Ultrasonic wire bonding quality monitoring using piezoelectric sensor *Sensors and Actuators A* **65** 69-75

Author

Wuwei Feng, born on January 19, 1980, Shanxi province, China

Current position, grades: associate professor of Zhejiang Ocean University.
University studies: Ph. D. degree in mechanical-electronics engineering from Xi'an Jiao tong University in 2011.
Scientific interest: microelectronic packaging quality control, digital signal processing, mechanical system fault diagnose.
Publications: 20 papers.

Micro-macroscopic statistical description on damage evolution of concrete

Ping Li[1*], Yongchi Li[1], Shiwei Duan[2], Mingtao Liu[1], Ruiyuan Huang[1]

[1]*Department of Modern Mechanics, University of Science and Technology of China, Hefei 230027, China*

[2]*School of Mechanical Engineering, Anhui University of Technology, Maanshan, 243000, China*

Abstract

In this paper, we present two types of damage model, coupled pressure-shear damage model and tensile damage model, according to the different stress state and different damage mechanism in concrete. Combining microscopic and macroscopic mechanics, we derive the damage evolution equation and obtain relevant material parameters by fitting the test data of one-dimensional compression and tension test. In order to verify the proposed damage model, we carry out numerical simulation on wave propagation problems caused by the explosive charge in concrete columns. The simulation results are consistent with experimental results, which show successfully the evolution of damage in explosion process.

Keywords: concrete, coupled pressure-shear damage, tensile damage, damage evolution

1 Introduction

It is well known that in brittle materials such as concrete, there are two kinds of damage forms, one is the tensile damage caused by tensile stress, and another is the coupled pressure-shear damage related to plastic volume dilatation phenomena caused by coupling effect of pressure and shear. From this point of view, in [1-4], damage is defined as a function of strain or stress to describe the material damage softening. And in [5], the damage is defined as a function of the hydrostatic pressure, equivalent plastic stain and accumulated plastic volumetric strain, which has been widely used in engineering and adopted by the commercial software LS-dyna. In reference [6], the damage evolution was related to the average tensile stress, crack density and volume strain, which can better reflect the cratering and caving in the penetration problem. Although many scholars have carried out a lot of work about the damage evolution of concrete and they have obtained some important achievements, there are still some shortages. First, most of the researches [7-11] are based on the macroscopic or microscopic damage theory, however, the macroscopic description has not combined together well with microscopic mechanism and hence the damage model lacks microscopic physics background; and in the microscopic model, it is difficult to combine macroscopic mechanical quantity with a variety of microscopic material parameters measured by precision instruments. Secondly, the different damage mechanism in different stress state has not been specifically considered.

Many studies such as Li Yongchi et al. [12] have demonstrated that there are mainly two different types of damage form for concrete-like materials: one type of damage behaves as growth and connection of micro-voids (supplemented by extension and connection of micro-crack), which is microscopically caused by the falling off and rupture of molecular bonds, and macroscopically shows the so-called "plastic volume dilatation phenomenon" related to pressure-shear coupling yield properties, that is the "coupled pressure-shear" damage. The other type of damage is the tensile damage which behaves as extension and connection of micro-crack (supplemented by expansion and connection of micro-void) caused by tensile stress in concrete. However, for current research on the damage of concrete, some damage models [13-15] are only suitable for simple stress state (tensile, compressive or shear). Others can be only applied to the case that the effects of pressure and shear stress in material is decoupling [16-18], which can also not well describe the actual damage evolution process or reflect "plastic volume dilatation phenomenon" for the material.

In this paper, the damage in concrete is divided into the coupled pressure-shear damage and tensile damage according to the different stress states and different evolution mechanism. Combining microscopic and macroscopic mechanics, we derived the damage evolution equation and obtained relevant material parameters by fitting the test data of one-dimensional compression and tension test. To verify the proposed damage model, we carried out numerical simulation on wave propagation

problems caused by the explosive charge in concrete columns. The simulation results show that the distribution and evolution about the coupled pressure-shear damage and the tensile damage are in good agreement with experimental results.

2 The coupled pressure-shear damage

From the perspective of the material microstructure, there are large amounts of micro-voids and micro-cracks with different shapes, sizes, and spatial orientation, which are the micro-damage in concrete material. the "coupled pressure-shear" damage behaves as growth and connection of micro-voids, which is microscopically caused by the falling off and rupture of molecular bonds, and macroscopically shows the so-called "plastic volume dilatation phenomenon" related to pressure-shear coupling yield properties, The growth of these micro-voids results in damage evolution. Macroscopically we define the damage D as follows:

$$D = \frac{V_d}{V}, \quad V_d = DV, \tag{1}$$

where V is the bulk volume of the representative element, V_d is the total volume of micro-voids. Differentiating Equation (1) with time t, we have:

$$\dot{V}_d = \dot{D}V + \dot{V}D. \tag{2}$$

For the coupled pressure-shear damage, if the damage mainly occurs at the plastic stage and to evolve with plastic volume dilatation caused by coupling effects of hydrostatic pressure and deviatory stress, it can be assumed that the relative growth rate of each micro-void is proportional to the plastic formation work rate; and if the damage is believed to occur in both elastic stage and plastic stage, it can be assumed that the relative growth rate of each micro-void is proportional to the total work rate. As an example for simplicity, the latter has been chosen. Then we have:

$$\dot{v}_d(i)/v_d(i) = a_1\dot{W}/W_B, \tag{3}$$

where $v_d(i)$ is the volume of the *i-th* micro-void, W is the formation work, a_1 is a material constant, and W_B may be taken as the formation work at the peak of stress-strain curves. By summing Equation (3) for all micro-voids, and according to component, we have:

$$a_1\dot{W}/W_B = \frac{\dot{v}_d(i)}{v_d(i)} = \frac{\sum \dot{v}_d(i)}{\sum v_d(i)} = \frac{\dot{V}_d}{V_d},$$

i.e.:

$$\dot{V}_d/V_d = a_1\dot{W}/W_B. \tag{4}$$

From Equations (1), (2) and (4), the following damage evolution equation is obtained:

$$\dot{D} = (a_1 D\dot{W} - D\frac{\dot{V}}{V})/W_B = (a_1 D\dot{W} - D\frac{\dot{v}}{v})/W_B, \tag{5}$$

where v is specific volume. Equation (5) is the general form of the coupled pressure-shear damage evolution equation. If it is further assumed that the volume of the solid part $V_s = V - V_d$ is incompressible, i.e. $\dot{V}_s = 0$, by Equation (1) we will have

$$\dot{D} = (1-D)\frac{\dot{V}}{V}, \quad D\frac{\dot{V}}{V} = \frac{D\dot{D}}{1-D}. \tag{6}$$

Then Equation (5) can be transformed into the following damage evolution equation:

$$\dot{D} = a_1 D(1-D)\dot{W}/W_B. \tag{7}$$

Our task is to optimize the material parameters a_1 and W_B by numerical fitting the MTS experimental stress-strain curve. For this paper, we assume that the material obeys the following nonlinear constitutive relation with damage:

$$\sigma = E\varepsilon(1+b\varepsilon)(1-D), \tag{8}$$

where E is Young modulus, b is a dimensionless parameter which is to be determined together with a_1 and W_B in Equation (7) by numerical fitting the MTS experimental stress-strain curve. From Equations (7) and (8), for the one-dimensional compression experiment with constant strain rate $\dot{\varepsilon}_c$, we can obtain the following system of ordinary differential equations for σ, D, ε:

$$\begin{cases} \dot{\varepsilon} = \dot{\varepsilon}_c \\ \dot{\sigma} = E(1-D)(1+2b\varepsilon)\dot{\varepsilon} - E\varepsilon(1+b\varepsilon)\dot{D}, \\ \dot{D} = a_1 D(1-D)\sigma\dot{\varepsilon}/W_B \end{cases} \tag{9}$$

or

$$\begin{cases} \dot{\varepsilon} = \dot{\varepsilon}_c \\ \dot{\sigma} = E(1-D)(1+2b\varepsilon)\dot{\varepsilon}_c - E\varepsilon(1+b\varepsilon)a_1 D(1-D)\sigma\dot{\varepsilon}_c/W_B \\ \dot{D} = a_1 D(1-D)\sigma\dot{\varepsilon}_c/W_B \end{cases} \tag{10}$$

The initial conditions for Equations (9) and (10) are: $\varepsilon(t=0) = 0$, $\sigma(t=0) = 0$, and $D(t=0) = D_0$. The terminal condition is that the stress-strain curve will reach the end point when the material completely fails, and at this point, the damage reaches the ultimate damage D_c. The initial damage D_0 and the ultimate damage D_c will be determined by numerical fitting the experimental curves together with a_1, W_B and b.

The representative stress-strain curves of C40 concrete under the three kinds of constant strain rates obtained by

Li Ping, Li Yongchi, Duan Shiwei, Liu Mingtao, Huang Ruiyuan

MTS one-dimensional experiment are shown in Figure 1. The optimal material parameters are shown in Table 1, which are got by fitting the constant strain rate curve $\dot{\varepsilon} = \dot{\varepsilon}_c = 10^{-4} / s$. The comparison between experimental curve and simulate curve is shown in Figure 2, in which the relationship $D \sim \varepsilon$ between damage and strain is also given.
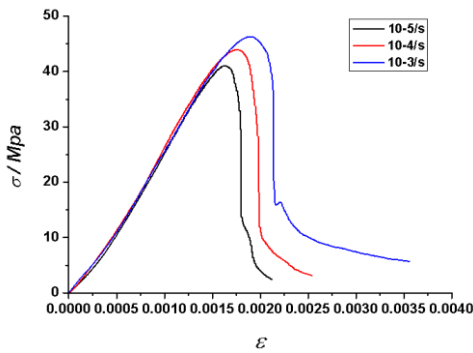
FIGURE 1 The constant strain rate stress-strain curves
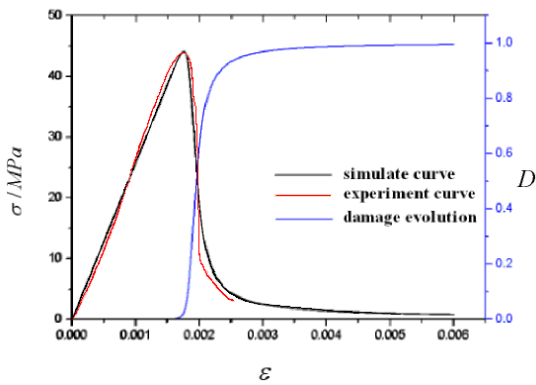
FIGURE 2 The stress-strain and damage-strain curves

TABLE 1 The material parameters of coupled pressure-shear damage evolution equation

$\dot{\varepsilon}_c$	D_0	a_1	D_c	b	W_b (MPa)
10^{-4}/s	10^{-5}	9.2	0.38	0.0001	0.04

3 The tensile damage

The main form of the tensile damage for brittle materials like concrete is the extension and connection of micro-crack caused by tensile stress in concrete. Taking mode I crack as an example, when the material is subjected to tensile stress mainly along a certain direction, the micro-cracks perpendicular to the stress will become the dominant growing crack which is the main reason for the fracture of brittle material.

Imagine there is a thin plate with a width L and unit thickness, in which there is a series of micro cracks with length of l_i perpendicular to the stress inside the material. The micro crack in representative volume element is shown in Figure 3.
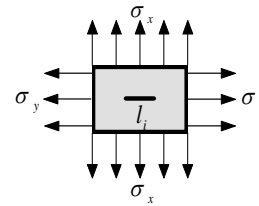
FIGURE 3 The micro crack of representative volume element

The macroscopic damage D is defined as a ratio of the total effective length of micro-crack and the total material length. The damage is assumed to be irreversible:

$$D = \frac{L_d}{L} = \frac{L - L_s}{L}, \quad \dot{D} \geq 0 , \qquad (11)$$

where L_d is the total length of micro cracks, L_s the total length of no damage solid material within the total length L of material, $L_s = L - L_d$. From Equation (11) we can obtain:

$$\dot{L}_d = \dot{D}L + D\dot{L} . \tag{12}$$

On microscopic scale, a large number of micro-cracks with the different size inside the material can be studied with the statistical methods, the growth of each crack may be assumed statistically independent, the interaction between micro-cracks is ignored; Thus the total length L_d of the crack in material can be described by the statistical sum of all the micro-cracks length in material, that is:

$$L_d = \sum l_i . \tag{13}$$

In order to analyse the growth process of a single micro-crack, the following hypothesis similar to Fen et al. [19] for micro void is made.

1) In the process of micro cracks growing up, the surface energy due to an increase in surface area of the micro-crack is supplied by the elastic strain energy of a limited area around the medium, and the limited area is determined by the velocity of Rayleigh wave propagation along the crack surface.

2) The threshold condition of micro-cracks growing up can be determined based on Griffith fracture theory.

The threshold stress σ_c in critical state for quasi-static growth of the plane stress crack (length 2l) is:

$$\sigma_c = \sqrt{\frac{2\eta \mathrm{E}}{\pi l}} , \tag{14}$$

where E is the Young modulus, and η is the surface energy per unit area. The strain energy of unit volume is:

$$w = \int \sigma d\varepsilon = \int_{\sigma_c}^{\sigma} \frac{1-\nu^2}{E} \sigma d\sigma = \frac{1-\nu^2}{2E}\left(\sigma^2 - \sigma_c^2\right), \qquad (15)$$

where ν is Poisson ratio. In the process of micro-cracks (length $2l$) growing up, the elastic strain energy in volume element is:

$$W = \frac{1-\nu^2}{2E}\left(\sigma^2 - \sigma_c^2\right)2l \cdot 2l \cdot 1 = \frac{2\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)}{E}l^2 . (16)$$

After the crack (length $2l$) growing $2\Delta l$, the increase of surface energy ΔU in volume element is:

$$\Delta U = 4\eta\Delta l . \qquad (17)$$

The released strain energy ΔW of limited area around the medium can be obtained by Equation (16)

$$\Delta W = \frac{4\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)}{E}lC_R\Delta t , \qquad (18)$$

where C_R is the velocity of Rayleigh wave, which is equal to:

$$C_R = \frac{0.862+1.14\nu}{1+\nu}C_s = \frac{0.862+1.14\nu}{1+\nu}\sqrt{\frac{G}{\rho}} = \frac{0.862+1.14\nu}{1+\nu}\sqrt{\frac{E}{2\left(1+\nu\right)\rho}} , \qquad (19)$$

where C_s is the velocity of transverse wave, G the shear modulus, ρ the density, ν the Poisson ratio.

According to the above hypothesis (1), $\Delta W = \Delta U$, we have:

$$\Delta l = \frac{\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)}{\eta E}lC_R\Delta t . \qquad (20)$$

So the growth rates of a single crack is:

$$\dot{l} = \frac{\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)}{\eta E}lC_R . \qquad (21)$$

And using Equations (13) and (21), the rate of change of the total length of the crack is:

$$\dot{L}_d = \frac{\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)C_R}{\eta E}L_d . \qquad (22)$$

This is the kinetic equation of crack growth. Using Equations (12) and (22) we can get:

$$\dot{D} = \frac{1-\nu^2}{\eta E}(\sigma^2 - \sigma_c^2)C_R D - \frac{D}{L}\dot{L} . \qquad (23)$$

Equation (23) is the general form of tensile damage evolution equation.

If the growth length of crack is assumed to equal the decrease length of the solid material, namely $\dot{L}_d \approx -\dot{L}_s$, then $\dot{L} = \dot{L}_d + \dot{L}_s = 0$ and Equation (23) can be simplified as:

$$\dot{D} = \frac{\left(1-\nu^2\right)\left(\sigma^2 - \sigma_c^2\right)C_R}{\eta E}D . \qquad (24)$$

The condition of damage developing is $\sigma \geq \sigma_c$, the tensile stress exceeds the threshold stress σ_c.

For our concrete material C40, the tensile fracture strength by experiment is $f_t = 3.48MPa$, which will be taken as the threshold stress σ_c, namely $\sigma_c = f_t = 3.48MPa$. The Young modulus is $E = 29.3GPa$ and the Poisson's ratio is $\nu = 0.20$. The specific experimental methods and specimen preparation have been illustrated in Li Ping et al. [20]. When the micro-crack length is taken as $130\mu m$, the surface energy η will be 0.08 J/m² by Equation (14). And the velocity of Rayleigh wave C_R obtained using Equation (19) is:

$$C_R = \frac{0.862+1.14\nu}{1+\nu}\sqrt{\frac{E}{2\left(1+\nu\right)\rho}} = \frac{0.862+1.14\times0.2}{1+0.2}\sqrt{\frac{29.3GPa}{2\left(1+0.2\right)2306kg/m^3}} = 2109m/s .$$

The material parameters in tensile damage evolution are shown in Table 2.

TABLE 2 The material parameters in tensile damage evolution

ν	$\eta(J/m^2)$	$E(GPa)$	$\sigma_c(MPa)$	$C_R(m/s)$
0.2	0.08	29.3	3.48	2109

4 The wave propagation problems caused by explosive load

As an example to verify and apply the damage models presented, a numerical simulation of the cylindrical concrete under the internal uncoupled explosive load has been carried out by the finite element program HVP similar to EPIC-2 [21, 22]. By the experiments on MTS with confining pressure and SHPB, we have obtained the following dynamic coupled pressure-shear viscoplastic yield criterion for damaged concrete:

$$\bar{\sigma} = f_c\Psi(p^*)\left(1 + B\ln\left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_0}\right)\right)(1-D) , \qquad (25)$$

where:

Li Ping, Li Yongchi, Duan Shiwei, Liu Mingtao, Huang Ruiyuan

$$\Psi\left(p^{*}\right) = \sigma_{m}^{*} - \gamma f_{c} e^{-\beta(p+p_{3t}^{*})/f_{c}} ,$$

$$p^{*} = \frac{p}{f_{c}} , \quad p_{3t}^{*} = \frac{p_{3t}}{f_{c}} , \quad \sigma_{m}^{*} = \frac{\sigma_{m}}{f_{c}} . \tag{26}$$

where $\bar{\sigma}$ is Mises equivalent stress and f_{c} is the static uniaxial compression strength, variables with the superscript "*" are dimensionless quantities normalized by the static uniaxial compression strength f_{c}; the influence of hydrostatic pressure on the yield strength is described by the yield factor $\Psi(p^{*})$ which is a function of hydrostatic pressure, $p^{*} = \dfrac{p}{f_{c}}$ is the normalized pressure p is the actual pressure, p_{3t} is the maximum hydrostatic tensile strength, $\sigma_{m}, \gamma, \beta$ are the material parameters; $\dot{\varepsilon}$ is the strain rate, and $\dot{\varepsilon}_{0}$ is the reference strain rate, B is the strain rate coefficient; $D = D_{1} + D_{2}$ is the total damage which includes the coupled pressure-shear damage D_{1} and the tensile damage D_{2}, i.e. the two kinds of damage will all decrease the material strength with a linearly weakening manner. By experiments of triaxial confining pressure, we have also obtained the relationship between the hydrostatic pressure and volume strain η:

$$p = k_{0}\eta + k_{1}\eta^{2} + k_{2}\eta^{3} + k_{3}\eta^{4}, (\eta = v / v_{0} - 1), \tag{27}$$

where k_{0}, k_{1}, k_{2} and k_{3} are material parameters. All relevant material parameters in Equations (25)-(27) are shown in Table 3

TABLE 3 The material parameters of concrete

k_0 (GPa)	k_1 (GPa)	k_2 (GPa)	k_3 GPa
6.98	124.50	-3251.3	21330
γ	β	p_{3t}^{*}	σ_{m}^{*}
6.7825	0.2958	0.025	7.0

The calculation model is shown in Figure 4: the outer radius of concrete column is 30 cm, the inner radius is 5 cm, the radius of TNT explosive in the center of concrete column is 1 cm, the foam is padded between the explosive and concrete materials. The explosive column is detonated by means of centerline initiation.
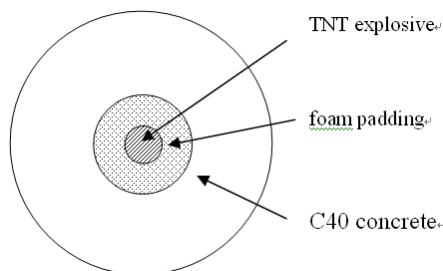
FIGURE 4 The calculation model

The calculation results are shown in Figures 5-8, in which Figures 5 and 6 are the moiré patterns of coupled pressure-shear damage at the *t*=45 μs and *t*=90 μs after the centre initiation respectively. It can be seen from the two figures that in the inner region of the concrete column, the coupled pressure-shear damage reaches the ultimate value accompanying with completely crushing destruction of material; while in the outer region of the concrete column, the coupled pressure-shear damage has not yet occurred, and in the middle region of concrete column, the damage value is between initial damage and ultimate damage. Due to the probability perturbation method adopted in the calculation, the damage manifests itself an asymmetric distribution in the circumferential direction, which is consistent with experimental results.
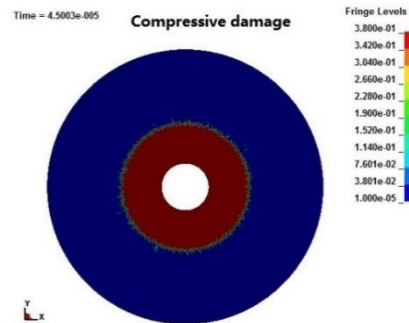
FIGURE 5 *t*=45μs coupled pressure-shear damage moiré pattern
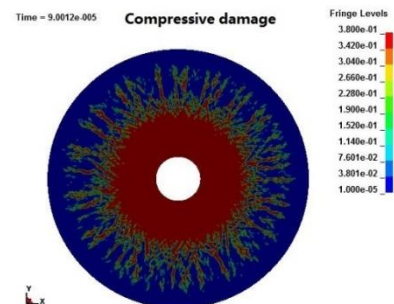
FIGURE 6 *t*=90μs coupled pressure-shear damage moiré pattern

Figures 7 and 8 are the moiré patterns of tensile damage at the *t*=45 μs and *t*=90 μs after the center initiation respectively. It can be seen from the two figures that in the inner region of the concrete column, there is no tensile damage occurring; while in the outer region, the tensile damage occurs because of tensile stress exceeding threshold σ_{c} generated due to the reflection of compression wave from the free surface of concrete column. As the case for coupled *c* pressure-shear damage calculation, the probability perturbation method is also adopted in the tensile damage calculation, and the tensile damage also manifests itself an asymmetric distribution in the circumferential direction.

Li Ping, Li Yongchi, Duan Shiwei, Liu Mingtao, Huang Ruiyuan
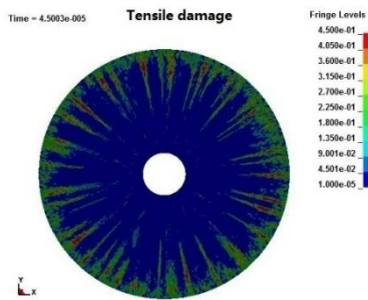
FIGURE 7 *t*=45μs tensile damage moiré pattern
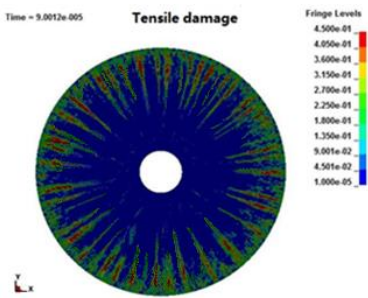
FIGURE 8 *t*=90μs tensile damage moiré pattern

The above calculation results are reasonable and credible by compared with experiment [23] and the theoretical analysis, and verifies the dynamic constitutive model with damage put forward in the paper is scientific and practical.

5 Conclusions

1) In this study, the damage in concrete is divided into the coupled pressure-shear damage and tensile damage according to the different stress states and different evolution mechanism;

2) Based on the idea "concrete being a system with many micro voids" and "growth of existent damage-nucleus in material", the evolution equation of the coupled pressure-shear damage is obtained by the microscopic statistics method. The relevant material parameters in the evolution equation are determined by numerical fitting the test data of one-dimensional compression.

3) Based on the idea "concrete being a system with many micro cracks" and "the crack surface energy is supplied by the elastic strain energy around cracks", the evolution equation of the tensile damage is obtained by the microscopic statistics method. The relevant material parameters in the evolution equation are determined from concerned theory and some test data.

4) To verify the proposed damage models, numerical simulation on wave propagation problems caused by the uncoupled explosive charge in concrete columns has been made. The simulation results show that the distribution and evolution about the coupled pressure-shear damage and the tensile damage are in good agreement with experimental results.

References

[1] Xie H, Sanderson D J 1995 Fractal kinematics of crack propagation in geomaterials *Engineering Fracture Mechanics* **50**(4) 529-36

[2] Shen X, Shen G, Chen L 2004 Study on a plastic constitutive damage model for concrete *Rock and Soil Mechanics* **25**(2) 13-6 *(in Chinese)*

[3] Lee J, Fenves G L 1998 Plastic-damage model for cyclic loading of concrete structures *Journal of Engineering Mechanics* **124**(8) 892-900

[4] Xie H 1995 Effect of fractal crack *Theory & Application of Fractal Mechanism* **23** 235-44

[5] Holmquist T J, Johnson G R, Cook W H 1993 A computational constitutive model for concrete subjected to large strains, high strain rate, and high pressures *The 14ᵗʰ International Symposium on Ballistics* Quebec City Canada 1993 591-600

[6] Taylor L M, Chen E P, Kuszmaul J S 1986 Microcrack-induced damage accumulation in brittle rock under dynamic loading *Journal of Computer Methods in Applied Mechanics and Engineering* **55**(3) 301-20

[7] Comi C, Perego U 2001 Fracture energy based bi-dissipative damage model for concrete *International Journal of Solids and Structures* **38** 6427-54

[8] Krajcinovic D, Fanella D 1986 A micromechanical damage model for concrete *Engineering Fracture Mechanics* **25**(5) 585-96

[9] Salari M R, Saeb S, Willam K J, Patchet S J, Carrasco R C 2004 A coupled elastoplastic damage model for geomaterials *Computer Methods in Applied Mechanics and Engineering* **193**(27) 2625-43

[10] Grassl P, Jirásek M 2006 Damage-plastic model for concrete failure *International Journal of Solids and Structures* **43**(22) 7166-96

[11] Voyiadjis G Z, Taqieddin Z N, Kattan P I 2008 Anisotropic damage–plasticity model for concrete *International Journal of Plasticity* **24**(10) 1946-65

[12] Li Y 2012 Introduction to Tensor Analysis and Modern Continuum Mechanics *Hefei: University of Science and Technology of China Press (in Chinese)*

[13] Zhang Z 2003 Dynamic damage and failure model for concrete materials against projectile penetration *Explosion and Shock Waves* **23**(1) 67-8 *(in Chinese)*

[14] Riedel W, Thoma K, Hiermaier S, et al. 1999 Penetration of reinforced concrete by BETA2B2500 numerical analysis using a new macroscopic concrete model for hydrocodes *The 9th International Symposium Interaction of the Effects of Munitions with Structures* Berlin-Strausberg: IBMAC 1999 315-22

[15] Ma A, Huang F, Chu Z, Li J 2008 Numerical simulation on yawed penetration into concrete *Explosion and Shock Waves* **28**(1) 33-7

[16] Shao C, Qian Y 2010 A kind of scalar damage elastoplastic constitutive model of concrete *Journal of Computational Mechanics* **27**(2) 298-302

[17] Burlion N, Gatuingt F, Pijaudier-Cabot G, Daudeville L 2000 Compaction and tensile damage in concrete: constitutive modeling and application to dynamics *Computer Methods in Applied Mechanics and Engineering* **183**(34) 291-308

[18] Jin Q 2006 Dynamic damage and failure model for concrete materials *Acta Armamentarii* **27**(1) 10-4

[19] Fen J 1992 Damage function model of dynamic ductile fracture of metals *Beijing Institute of Technology (in Chinese)*

[20] Li P, Li Y, Duan S, and Huang R 2013 A contrast experimental study of C40 concrete tensile strength. *Advanced Materials Research* **815** 700-6

[21] Johnson G R 1976 Analysis of elastic-plastic impact involving severe distortions *Journal of Applied Mechanics ASME* 439-44

[22] Johnson G R, Stryk R A 1986 User instructions for the EPIC-2 code *AFATL-TR-86-51 Honeywell Defense Systems Division*, Edina, MN

[23] Zhang H, Duan Z, Liu Y 2013 Study on the collapse perforation of thick concrete targets under Internal Explosion *Transactions of Beijing Institute of Technology* **33**(5) 441-44

Authors

Ping Li, born in December, 1985, Shandong province, China

Current position, grades: doctor student at University of Science and Technology of China.
University study: Master's degree in University of Science and Technology of China, 2011.
Scientific interests: Computer modelling and simulation of the materials, mechanics modelling, and material impact dynamics analysis.
Publications: 2 papers.
Experience: 1 project, key projects of national natural science funds.

Yongchi Li, born in December, 1941, Hebei province, China

Current position, grades: professor at the University of Science and Technology of China.
University study: PhD degree in Beijing university, 1965.
Scientific interests: Impact mechanics numerical method and simulation calculation of various kinds of engineering problems, finite difference, finite element theory, feature of the new calculation method research and software development.
Publications: 200 papers, 1 book.
Experience: Doctoral supervisor in the university of science and technology of China, 1990.

Shiwei Duan, born in August, 1984, Hebei province, China

Current position, grades: doctor at the University of Science and Technology of China. Associate professor in Anhui University of Technology.
University study: PhD degree in University of Science and Technology of China, 2013.
Scientific interests: computer modelling and simulation of material, engineering structure reliability analysis, mechanics modelling.
Publications: 3 papers.
Experience: key projects of national natural science funds.

Mingtao Liu, born in February, 1986, Shanxi province, China

Current position, grades: Doctor student at the University of Science and Technology of China. Researcher at the China academy of physics.
University study: Master's degree in University of Science and Technology of China, 2011.
Scientific interests: Computer modeling and simulation of the materials, engineering structure reliability analysis, engineering technology
Publications: 3 papers.

Ruiyuan Huan, born in February, 1984, Fujian province, China

Current position, grades: doctor at the University of Science and Technology of China.
University study: PhD degree in University of Science and Technology of China, 2013.
Scientific interest: the dynamic performance of the materials and structure, mechanics modelling, structural safety analysis.
Publications: 3 papers.
Experience: project from national postdoctoral fund.

Numerical simulation for mine rescue capsule gas explosion dynamic response

Jianying Liu[1, 2*], Qiaoxin Zhang[1], Lixia Deng[2]

[1]*Wuhan University of Technology, Wuhan, Hubei, 430070, China*

[2]*Henan Engineering College, Zhengzhou, Henan, 451191, China*

Received 2 August 2014, www.cmnt.lv

Abstract

In order to ensure the research and development production of coal mine "KJYF - 96/8 portable hardware capsule" in the coal mine use can effectively resist gas explosion shock pressure, structural strength meet the safety requirements. The finite element model is set up, to take specific gas coal dust explosion triangle under the action of shock waves, the stress, strain and displacement response. The results show that the peak of 0.6 MPa and 1.2 MPa, 300ms triangle shock wave, the capsule basic no plastic deformation, overall no failure parts, ships were able to be whole, the structure safety. Depending on the overall displacement nephogram view of displacement nephogram and components parts after shock wave action, maintain the overall coordination deformation, minimal relative displacement, sealing well.

Keywords: capsule, gas explosion, shock wave, dynamic response, numerical simulation

1 Sample description and explanation

1.1 THE CAPSULE GAS MAIN STRUCTURE, MATERIAL, SIZE

KJYF-96/8 mine mobile hardware lifesaving cabin assembled into a split. The body is divided into four parts from the structure: transition cabin, cabin escape pod, personnel, equipment cabin, transition cabin crew capsule 1.2 meters, 3.6 meters, 1 meter of escape pod, equipment cabin crew capsule 3.9 meters, are provided with personnel doors and maintenance hatch. The escape pod length is 1 meter, the escape pod unilateral with escape door, equipment cabin length is 3.9 meters, the size structure.

Intermediate basic cabin unit is whole steel structure, both ends are flange structure, the basic unit of the adjacent cabin flange connection structure. Capsule thickness is 12 mm basic cabin, cabin bolted flange thickness of 25 mm. The basic bodies through flange structure are connected by bolts. Both ends of the flange plate and the plate adopt welding connection. When assembling the internode with M16 bolts, internodes with seal. The front door plate is in thickness of 25 mm. The main doors hole size is 600 mm×1200 mm, 420 × 670 mm emergency escape door. Observation window for explosion-proof glass is 70 mm, materials for the FBZ-B-70-HX. Reinforcing rib is arranged outside the cabin, 5 channels, reinforced chassis is provided with 8 channels consisting of slipper, chassis is provided with 10 channels. The door is connected by a shaft seat and the end shaft. Lifesaving cabin chassis slipper and slide placed with [1-4].

1.2 THE CAPSULE GAS FIXED WITH THE GROUND

Chassis slipper of KJYF-96/8 coal mine mobile hardware capsule is provided with 10 steel compositions, capsule chassis slipper and slide placed with. It can be capsule and the tunnel bottom surface constraint simplification is determined according to the actual size, simply supported connection.

1.3 THE ANTIKNOCK STRESS ANALYSIS REQUIREMENTS

The peak of 0.6 MPa, for 300 ms gas and coal dust explosion triangle shock wave; limit peak triangular shock reaching overall structural damage, damage of the key parts of 1.2 MPa wave.

2 The numerical calculation

2.1 THE CALCULATION SOFTWARE

Calculation under shock wave loading KJYF-96/8 mine mobile hardware lifesaving cabin stress and plastic deformation, large displacement problem is non-linear, time-varying shock wave loading. The explicit nonlinear dynamic analysis program ANSYS/LS-DYNA for numerical simulation, mechanical model of cabin under shock wave response.

Corresponding author e-mail: liujianying74@163.com

2.2 CALCULATION MODEL

2.2.1 The choice of unit type

The KJYF-96/8 mine mobile hardware capsule simulation antiknock performance numerical modeling on the basis of the actual need, the size, modeling preserved during the main structure characteristics, reasonable simplification for small parts. Reasonable selection of unit types, to shorten the time for solving the model for improving the precision of simulation, the effective rescue capsule plays an important role in the shift of shock wave. ANSYS/LS-DYNA in the application display may use in the dynamic analysis of unit: bar element, beam element BEAM161, LINK160 surface unit PLANE162, SHELL163 thin shell element, solid element SOLID164, spring damper element COMBI165, MASS166 and other quality unit.

Simulation by finite element under impact load values of the rescue capsule, the selecting unit, to consider the need to model the actual size and calculation, the lifesaving cabin basic cabin skin thickness relative to the rescue capsule size thickness is very thin, so the basic hull skin plate shell element SHELL163 mesh is reasonable. Need to door key processing, according to the size, determine the entity unit SOLID164 divided the main doors and doors mesh is more reasonable. The solid elements SOLID164 flange structure grid division. Bolt connection, using spring damper element COMBI165 mesh [5-9].

2.2.2 The mesh

The structural characteristics of cabin model, the shape is irregular, the different combinations of parameters after many experiments, a general parameters and the mesh parameter: ratio coefficient is the default value of 1, set the maximum cell size of 50 mm shell element, solid element

maximum element size is 25 mm, the details at the maximum unit size is 10~15 mm, is divided into thin shell element, solid element, rigid unit 429545, the final model selected grid more appropriate parameters. KJYF-96/8 mine mobile hardware lifesaving cabin model grid, see Figure 1. The finite element model of node and element number is moderate, the lifesaving cabin loading explosive shock wave calculation accuracy. The grid quality check, cell size is uniform, meet the engineering requirements.

FIGURE 1 KJYF-96/8 mine mobile hardware capsule mesh model

3 Material model

Nonlinear material constitutive relation are elastoplastic finite element method, the rigid plastic finite element method, elastic viscoplastic finite element method, the rigid viscoplastic finite element method four types, as shown in Figure 2. Metal material constitutive relationship can reflect the accuracy of material properties of metal with finite element simulation. The capsule will produce plastic deformation under shock wave, the main concern of plastic deformation, therefore, lifesaving cabin in the wave finite element analysis using LS-DYNA Johnson-Cook in shock when the strain rate effect and the effect of temperature on the elastic-plastic constitutive model considering.
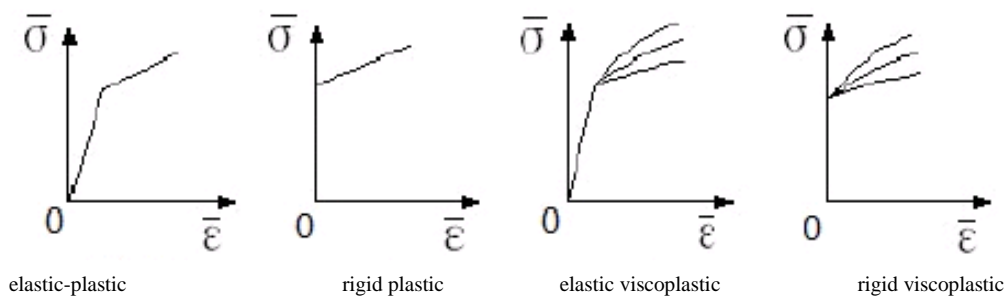
FIGURE 2 Metal material nonlinear constitutive curves

4 The initial conditions and boundary conditions

Because the KJYF-96/8 coal mine mobile hardware lifesaving cabin structure mainly by welding and bolt connecting part, not the relative movement of the parts. So the hypothesis: welding structure is completely reliable, structural parts have penetration, there is no welding stress, welding leg length of structure has no effect; the capsule

structure does not exist any the manufacture or installation deformation; bolt connection is reliable, it has no effect on the structure of pre-stressed.

4.1 BOUNDARY CONDITIONS

In the ANSYS implicit finite element analysis program, the contact interaction between moving objects, the

contact element to simulate, it makes the analysis process is extremely complex and very difficult to understand. In the ANSYS/LS-DYNA program, no contact element, as long as the definition may contact surface, the contact type and contact with some related parameters, the contact interface does not occur between the penetration guaranteed in the calculation process, and consider the effect of friction at the contact interface for relative motion. Contact types include single contact (Single Surface), the node to surface contact (Nodes to Surface) and surface to surface contact (Surface to Surface) several. The simulation of blast wave to the rescue capsule, directly to the lifesaving cabin pressure, there is no definition of contact problems.

Chassis slipper of KJYF-96/8 coal mine mobile hardware chassis is provided with 10 steel compositions, lifesaving cabin chassis slipper and slide placed with. It can be lifesaving cabin and the tunnel bottom surface constraint is modeled as a simply supported is determined according to the actual size of the connection.

4.2 THE INITIAL CONDITIONS

After a simple model analysis showed that, lifesaving cabin elastoplastic deformation under explosive blast wave, deformation mainly occurred in the capsule and blast shock wave contacts the outer boundary, the deformation amplitude relative to the rescue capsule size is very small, the reasonable mesh, mesh deformation is not too serious, in view of the Lagrange algorithm to deal with the free surface and material the interface intuitive and natural, and can track material boundary accurately and describe the material interface and other advantages, the Lagrange algorithm. The initial conditions of static, in addition to the capsule itself gravity, do not consider other load. Method there is usually several create shock wave pressure, one is to build a preset pressure time curve, and the other is the numerical simulation of explosive effect. Generated by explosive explosion pressure pulse in the air grid, will limit the pressure pulse amplitude, duration and other parameters. Considering the calculation procedure and the actual effect of explosion shock wave, calculated using the method of loading the pressure wave simulation of shock wave on the class action.

The lifesaving cabin under explosive blast wave may damage into consideration, take lifesaving cabin front, back end face is positive, according to the peak load; the rest of the surface according to half peak. Due to the presence of certain errors between the numerical simulation calculation and real situation, and it is necessary to consider the safety factor, the calculation of the shock wave overpressure is tentatively scheduled for 0.6 MPa, peak value and pressure of explosion shock wave 1.2 MPa calculations.

5 Hull structure response calculation results and analysis

At 0 ms, forward propagating shock wave reaches the capsule surface, the KJYF-96/8 mine mobile hardware lifesaving cabin applied peak for 0.6 MPa, 1.2 MPa for 300 ms, triangle shock wave pressure, the computation time is 400 ms. Read the waves on the lifesaving cabin effect of explosion, achieve the deformation results, nephogram rendering nodes or unit, the interception of the required information.

Two kinds of loads, lifesaving cabin structure stress and displacement nephogram because of the limited space omission but strain, shown in Figure 3.
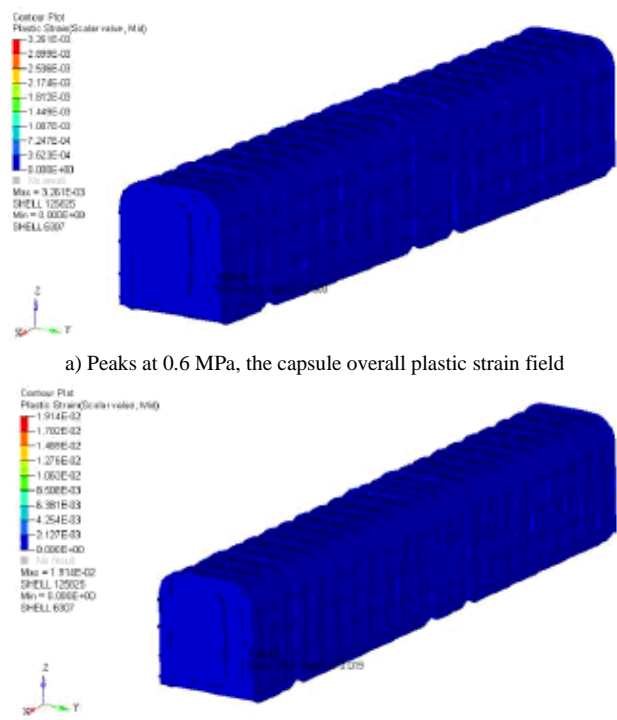
a) Peaks at 0.6 MPa, the capsule overall plastic strain field

b) Peaks at 1.2 MPa, the capsule overall plastic strain field
FIGURE 3 Capsule overall plastic strain field

The lifesaving cabin overall typical nodes (unit) as an example, the KJYF-96/8 mine mobile hardware lifesaving cabin in the coal mine gas explosion under conditions of stress and displacement time history.

1) Capsule whole stress maximum unit stress time history results.

Two kinds of loads, the stress reaches the maximum value at the unit 125825th unit and 2089th unit, extraction of σ-t curve (stress time curve), shown in Figure 4.

Visible, the lifesaving cabin under the wave crest is 0.6 MPa and 1.2M Pa two kinds of load, stress values were 0~260.0 MPa, 0~363.5 MPa, lifesaving cabin to meet the strength requirements.

a) Peaks at 0.6 MPa, the capsule stress peak position unit σ-t curve

b) Peaks at 1.2 MPa, the capsule stress peak position unit σ-t curve

FIGURE 4 Capsule stress peak positions unit σ-t curve

2) The displacement peak 42307 nodes and 42308 nodes, extract the delta time curve δ-t (displacement time curve), curve, shown in Figure 5.

Therefore, two kinds of loads, lifesaving cabin overall displacement values are 0~7.941 mm, 0~15.82 mm. Capsule is meeting the stiffness requirements, and no failure.

a) Peaks at 0.6 MPa, the capsule reach maximum displacement of the nodes of the δ-t curve

b) Peaks at 1.2 MPa, the capsule reach maximum displacement of the nodes of the δ-t curve

FIGURE 5 Capsule stress peak positions unit δ-t curve

6 Conclusion

To ensure that the R & D and production of the "KJYF-96/8 mine mobile hardware capsule" can effectively resist the impact pressure of gas explosion in coal mine use, structure strength to meet the safety requirements, using explicit nonlinear dynamic analysis program, the finite element model is established, more comprehensively and objectively shows the gas explosion accident occurred, bear peak is 0.6 MPa 300 ms, 1.2 MPa, duration of gas and coal dust explosion triangle waves, KJYF-96/8 mine mobile hardware lifesaving cabin generated stress, plastic strain, displacement response.

Crest 0.6 MPa, persistent 300 ms triangular shock wave, lifesaving cabin no plastic deformation without failure, the whole parts, cabin remain intact, structural safety. According to the overall displacement nephogram and parts displacement nephogram view, parts in the shock wave, keep the overall coordination of deformation, displacement minimum, seal intact.

Crest 1.2 MPa, persistent 300 ms triangular shock wave, lifesaving cabin produces only approximate to the plastic deformation in the flange, zero rib frame, shell structure of the minimal unit, no overall failure components, cabin remain intact, structural safety. According to the overall displacement nephogram and parts displacement nephogram view, parts in the shock wave, keep the overall coordination of deformation, displacement minimum, seal intact.

By the analysis of visible, capsule displacement maximum value appeared in the middle position of the two side capsule shell, such as the need to further improve the lifesaving cabin anti impact capability, can strengthen the internal reinforcement structure and quantity of the structure was strengthened, increase its flexural rigidity. Capsule stress reached the maximum at the connection position of the capsule shell and the flange frame structure, can strengthen the material and structure of these structures by local, to become further strengthen the anti-shock strength of the rescue capsule.

References

[1] Zhang Y, Yao B, Ye J 2007 Numerical simulation during shock wave propagation of gas explosion *Mechanical & Electrical Technology* **3** 28-30 *(in Chinese)*

[2] Wang S, Jing L, Li J 2010 The present status of overseas mine emergency refuge chamber technology *Journal of Safety Science and Technology* **4** 119-23 *(in Chinese)*

[3] Gao G, Zhang L 2009 Design principles of movable coal mine refuge chamber *Journal of Safety Science and Technology* **4** 162-4 *(in Chinese)*

[4] Cui X 2009 Mine removable life-saving capsule successfully developed *Shanxi Daily* **1** 4-17 *(in Chinese)*

[5] Zhuang Z 2009 The Finite Element Analysis and Application Based on ABAQUS *Beijing: Tsinghua University Press (in Chinese)*

[6] Zhang Y, Yao B, Ye J 2007 Numerical simulation during shock wave propagation of gas explosion *Mechanical & Electrical Technology* **3** 28-30*(in Chinese)*

[7] Margolis K A, Westerman C Y K, Trakofler K M K 2011 Underground mine refuge chamber expectations training: program development and evaluation *Safety Science* **49** 522-30

[8] Zhao H J, Qian X M, Li J 2012 Simulation analysis on structure safety of coal mine mobile refuge chamber under explosion load *Safety Science* **50** 674-8 *(in Chinese)*

[9] Fang H, Ge S, Cai L 2011 Buckling capacity optimization of coal mine refuge chamber's shell under uniform axialcompression *Proceedings of the 3rd International Conference on Measuring Technology and Mechatronics Automation Shanghai* 649-53

Authors

Jianying Liu, born in 1973, Shanxi, China

Current position, grades: associate professor of Henan Institute of Engineering.
University study: D.R. degree in mechatronic engineering, Wuhan University of Technology, Wuhan, Hubei, 2009.
Scientific interest: computer aided design and graphics technology.
Publications: 10 papers.

Qiaoxin Zhang, born in 1961, Hubei, China

Current position, grades: professor of Wuhan University of Technology.
University study: M.S. degree in mechatronic engineering, Huazhong University of Science and Technology, Wuhan, Hubei, 1988.
Scientific interest: the surface of the friction and wear, nano materials and nano sensor and industrial engineering.
Publications number or main: 60 papers.

Lixia Deng, born in 1972, Henan, China

Current position, grades: lecturer of Henan Institute of Engineering.
University study: M.S. degree in mechatronic engineering, Huazhong University of Science and Technology, Wuhan, Hubei, 2007.
Scientific interest: mathematics, electric control and PLC technology.
Publications number or main: 3 papers.

A disturbance detection interferometric fibre optic sensor system

Youneng Zhang¹*, Dawen Wang², Li Ding¹

¹*Department of electrical and information engineering of Anhui Industry &Trade Vocational Technical College, Huainan, Anhui, 232001, China*

²*Experiment center of Anhui Industry &Trade Vocational Technical College, Huainan, Anhui, 230001, China*

Abstract

This study proposed a Mach-Zehnder interferometer (M-Z)-based disturbance detection interferometric fibre optic sensor system. In this system, two single-mode optical fibres passing through the central tube of cable were used as the signal arm and reference arm of the M-Z interferometer. In the coherent length range of laser light source, as long as encountering suitable optical path difference, can the photosynthesis generate a series of alternately dark and bright interference fringes. Reference arm is dependent from the external signals, while the signal arm is subject to the disturbance of external signals. Affected by the disturbance, signal interference fringes significantly generated fluctuations with the variation of the optical distance between the two arms. The interference signals were then collected and stored through digital storage oscilloscope. By detecting the variations of the interference fringes from stability to instability, and from regularity to irregularity, it had access to judging whether or not there was external disturbance invasion. Experimental results suggested that the sensor system was very sensitive. In case of slight touch on the cable, the waveform will generate variations.

Keywords: interferometer, optical fibre senescing, disturbance, optical path difference

1 Introduction

Optical fibre can serve as the sensor in security protection. Once optical fibre sensor is influenced by external interferences, will part of characteristics in the light transmitted in the optical fibre generate variations. By analysing the signals collected by optical sensing devices and comparing the characteristic variations of the light detected, it is accessible to detecting and monitoring a number of invasion events and corresponding status. As for the optical fibre-based disturbance detection system, the intensity-modulated vibration sensor is advantageous in its simple structure, easily achievable optical path, and low production cost. However, it shows low precision and thus can merely used in the detection system with low requirements [1, 2]. Its working principle is indicated as that: by changing the phase difference of the two optical signals in Michelson interferometer [3], Mach-Zehnder interferometer [4-6], or Sagnac interferometer [7, 8] using vibration signals, the interference output changes with the sensing arm's length. According to the changes, the detection on external vibration signal detections is realized. However, such detection is conducted based on the internal structure of interferometer to trigger the variations of interference fringe when external signals reach to the interferometer. This is the traditional application of interferometer.

This paper proposed an M-Z interferometer-based new perturbation detection interferometric fibre optic sensor system. This system integrated the fibre in optical cable and the two arms of the M-Z interferometer into a new structural interferometer. On the basis of light interference theory, two single-mode fibres in the central tube of a 60 m cable were used as the two arms of the M-Z optical interferometer respectively. One arm was set as for reference, while the other one was used for sensing. Such application is difference with the transformation of general fibre interferometers. The fibre used as the reference arm is irrevent with the external signals and the optical phase delivered from this arm keeps constant, while the fibre serving as the sensing arm is subject to the disturbance of external signals. The variations of sensing arm are prone to cause the slight variations of the optical path. Thus in the coherent length of laser source, interference fringes generate significant variations. By detecting the variations of the interference fringes from stability to instability and from regularity to irregularity, it has access to judging whether or not there is external disturbance invasion.

2 The principle of the system

2.1 THE SENSING PRINCIPLE OF M-Z INTERFEROMETER

The sensing principle of Fibre M-Z interferometer is indicated as follows: by modulating the measured object, the optical properties of the measuring arm of interferometer, such as frequency and phase etc., generate

---

Corresponding author e-mail: 625491561@qq.com

variations. Since the modulated light is correlated with the local oscillator of the reference arm, intermediate frequency signals can be obtained through the frequency mixing by photoelectric detection components. The related information of the measured object was thereby perceived to realize the sensing function on the pre-measured information. Fibre M-Z interferometer is generally composed by single-mode fibre, two X or Y type couplers, light source, and photoelectric detection system [9]. Single-mode optical fibre, acting as the sensing element, receives and transmits the signals. Two optical fibre couplers are used as a multiplexer demultiplexer of light. As shown in Figure 1.
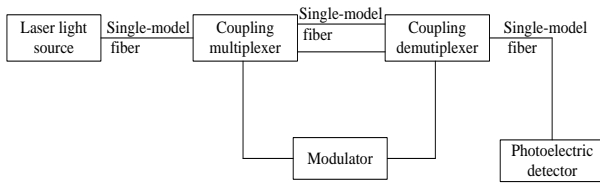
FIGURE1 The composed figure of M-Z interferometer

Being free of modulation, the optical transmission mode of the single-mode fibre is expressed as Equation (1) under the approximation of weak waveguide [10].

$$E_i(r) = \begin{cases} A'J_0\left(\dfrac{\mu}{r}\right), & 0 < r \le a \\[2mm] A'\dfrac{J_0(\mu)}{K_0(\mu)}K_0\left(\dfrac{\omega}{a}r\right), & r > a \end{cases} \quad (1)$$

where, $i=x$ or y expresses that the electric field component of the fibre polarize along direction x or y. a is the radius of the core of single-mode fibre,

$$\mu^2 = a^2(n_1^2 K^2 - \beta^2), \quad \omega^2 = a^2(\beta^2 - n_2^2 K^2); \quad k = \frac{2\pi}{\lambda};$$

n_1 and n_2 refers to the refection rates of the fibre core and outer layer respectively; $\beta = K_z$ is a longitudinal propagation constant; in the case of weak waveguide approximation, the longitudinal component E_z and H_z in optical fibre are very weak. Thus, the light in the fibre is approximated as the propagation of plane wave in the fibre. Considering the time factor of electric field component, the component is recorded as [10]:

$$E_i(r,t) = E_i(r)\cos(\omega_0 t + \varphi_0), \quad (2)$$

where, ω_0 and φ_0 refer to the angle frequency and initial phase of incident light respectively.

Under certain modulation, the maculated optical signal and the signals of the local oscillator in the other arm are combined by coupling demultiplexer and then irradiate to the surface of the photoelectric detector. In case of the electric vectors of the two beams being parallel, the total wave field is given as:

$$E_i(r,t) = E_{L0}(r)\cos(\omega_{L0} t + \varphi_{L0}) + E_{M0}(r)\cos(\omega_{L0} t + \varphi_{L0}), (3)$$

where, L_0 and M_0 represent the corresponding quantity of the local oscillator and modulated light respectively. The light current formed in the photocurrent photoelectric conversion can be expressed as:

$$i(t) = \frac{e^\eta P(t)}{h\gamma}, \quad (4)$$

η refers to the quantum efficiency of photoelectric detector. $P(t)$ is the efficiency of the light field irradiating to the surface of the detector:

$$P(t) = \frac{E^2(r,t)A}{Z_0}, \quad (5)$$

Z_0 is the wave impedance in the vacuum. A is the area of the surface detected. By combining Equations (2)-(5) is yielded:

$$i(t) = \frac{e\eta}{h\gamma}\frac{A}{Z_0}\frac{1}{2}E_{L0}\left[1 + \cos 2(\omega_{L0} t + \varphi_{L0})\right] + \frac{1}{2}E_{M0}^2\left[1 + \cos 2(\omega_{M0} t + \varphi_{M0})\right] + \frac{1}{2}E_{M0}^2\left[1 + \cos 2(\omega_{M0} t + \phi_{M0})\right] +$$
$$E_{L0}E_{M0}\cos\left[(\omega_{M0} - \omega_{L0})t + (\phi_{M0} - \phi_{L0})t\right] + E_{L0}E_{M0}\cos\left[(\omega_{M0} - \omega_{L0})t + (\varphi_{M0} - \varphi_{L0})t\right]. \quad (6)$$

Since ω_{M0}, ω_{L0}, and the combined frequency $\omega_{M0} + \omega_{L0}$ echo with the optical frequency, the detector shows no response and can be neglected. Thus Equation (6) transforms into Equation (7) finally [10].

$$i(t) = i_{L0} + i_{M0} + 2\sqrt{i_{L0}i_{M0}}\cos\left[(\omega_{M0} - \omega_{L0})t + (\varphi_{M0} - \varphi_{L0})t\right] = i_{L0} + i_{M0} + 2\sqrt{i_{L0} + i_{M0}}\cos(\omega_{if} t + \Delta\varphi), \quad (7)$$

where, i_{L0} and i_{M0} are DC terms and calculated by:

$$i_{L0} = \frac{e^\eta A}{2hYZ_0}E_{L0}^2, \quad (8)$$

$$i_{M0} = \frac{e^\eta A}{2hYZ_0}E_{M0}^2, \quad (9)$$

$\omega_{M0} - \omega_{L0} = \omega_{if}$ is intermediate frequency, $\omega_{M0} - \omega_{L0} = \Delta\varphi$ is initial phase difference. Since general modulated signal is far smaller than the optical frequency, there is $\omega_{M0} \ll \omega_{L0}$. Thus the light current varying with ω_{if} is detectable. It is given by:

$$i_{if} = 2\sqrt{i_{L0}i_{M0}}\cos(\omega_{if} t + \Delta\varphi). \quad (10)$$

By detecting the variations of the light current, it is able to perceive the variations of the pre-measured object.

2.2 THE SENSING PRINCIPLE OF THE DISTURBANCE DETECTION INTERFERMETRIC OPTICAL FIBRE

Figure 2 shows the sensing principle of the disturbance detection interfermetric optical fibre. The principle is indicated as that: basing on the M-Z interferometer, the two single-mode fibres in the central tube of a 60 m long cable are used as the reference arm and sensing arm of the M-Z interferometer respectively. Fibre coupler Y1 divides the light emitted from the laser source into two beams. Through the two single-mode fibres in the centre tube of the cable, the two beams were transmitted to the optical fibre coupler Y2. The optical fibre coupler Y2 combines the two beams into one and interfere the beam. The interfered beam develops into light current in the photo detector, Meanwhile, the optical signals are transformed into weak electric signals and then analysed under the amplification of the photoelectric detection circuit.

The disturbance detection interferometric fibre optic sensor system is based on M-Z interferometer. Since the interferometer arms are two single-mode fibres of a 60 m long optic cable, the light paths of the interferometer are basically symmetric. However, affected the uncertainties in the links of the production process of the two arms, such as the welding of light path and the processing of the end surface of fibre etc., the two arms are not completely symmetric in fact in length and show an error of 4 cm around. Attributing to this error, it is unable to guarantee that the symmetry of the two arms is in the coherent length of the light source. Therefore, the interference fringes produced in the disturbance detection system is not a yield of the coherence of the light source itself.
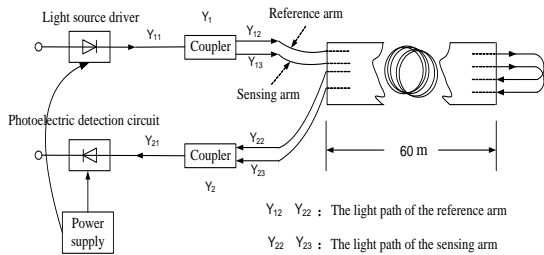
FIGURE 2 Schematic of disturbance detection interferometric fibre optic sensor system

Due to the objective existence of the length error of the two arms, there is light path difference between the two beams of the modulated light emitted from the laser divided by coupler Y_1. We set the lengths of the reference arm and sensing arms as L_1 and L_2 respectively, and the light path difference of the two arms as $\Delta L = c\Delta t$. The Δt therein refers to the time delay caused by the length difference of the two arms. Figure 3 shows the principle of production of the interference signals. The solid and dashed lines refer to the two beams to be interfered respectively. The ordinate represents the light frequency.

At time of $t = t_i$, the two beams show minor differences and develops into beat frequency signals. The beat frequency signals generated between t_{si} and t_{xi} keep constant and merely relates to the light path difference.
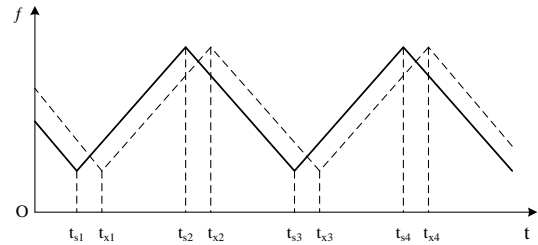
FIGURE 3 The Principle figure of interferometer singles

When the sensing arm is disturbed, the vibration frequencies of the two beams combined by fibre coupler Y_2 generate differences and optical "beat frequency". In the coherent length of the light source, as long as the optical paths of the two arms are suitable, it is prone to yield a series of alternately dark and bright fringes by photosynthesis. The sensing arm located between the two couplers. The fibre of reference arm is independent from the external signals and the optical phase therein keeps constant. When sensing arm fibre is subjected to the disturbance of external signals, the optical phase transmitted in the fibre will generate variations and produce optical phase difference in the case of the minor variations of light path, followed by the obvious variations of signal interference fringes. After the interference singles were then detected by the detection system and applied with signal processing, it is accessible to modulating the phase displacements of the pre-measured signals. When the reference light path and sensing light path are given the same disturbance simultaneously, the interference fringes of the signal basically remain constant. In case of stable disturbance, the interference fringes are supposed to be simple and harmonic vibrations under the influence of external forces, such as vibration [11]. When being applied with external disturbances, the cable generates symmetrical bending on a macro level. However, on the micro level (wavelength magnitude), the two single-mode sensing fibre are significantly asymmetric on bending and thus produce fringes (beat frequency) inevitably. The beat frequencies here, which are random variation frequency instead of regular harmonic frequency now, fluctuate randomly. Normally, the interference fringes vary slowly and stably. However, when there are external disturbance forces on the cable, the interference waves will show significant large fluctuations. Relying on these fluctuations, it is able to judge whether or not there are disturbances.

3 The structure of the system

Figure 4 shows the structure of the disturbance detection interfermetric fibre optical sensor system, which mainly includes the circuit part, light path part, and signal collection part.
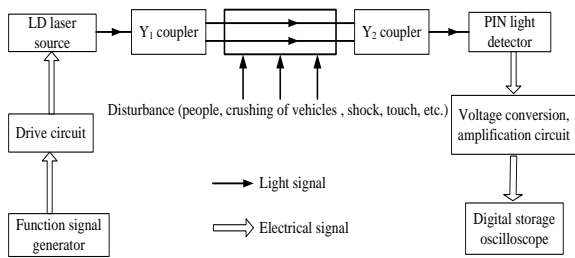
FIGURE 4 Schematic of diagram of experimental device of disturbance detection interferometric fibre optic sensor system

The triangle wave signals produced by the function signal generator are applied to the drive circuit and thus to input the triangle wave-formed current to the LD laser source. The output optical frequency of LD laser is the linear function of the input current. The optical carrier generated by LD laser source transformed the electric signals generated by the function signal generators into optical signals. Optical signals are divided into two beams after reaching to the coupler Y_1. The two beams are transmitted along the two fibres, namely, reference light path and sensing light path. Then the two beam of optical signals are integrated by coupler Y_2 into interference fringes. Since the sensing fibres in the disturbance detection interfermetric fibre optical sensor system are asymmetric in length, the two beams show time delay when reaching to the coupler Y_2 and develop into optical beat frequency.

After the sensing cables are allocated, the interference beat frequency signals is prone to yield significant variations in case of the external disturbances on the fibre. The interference lights are collected by the PIN detector. Related light signals are converted into electrical signals and finally displayed on the oscilloscope after voltage transformation and amplification. In case of no external disturbances, the interference waveforms are stored in the oscilloscope. By simulating the external disturbances, the waveforms under difference disturbances can be recorded for comparison and analysis.

4 Test Results

Figure 5 shows the experimental device. In the device, the waveforms were collected and stored by the digital storage oscilloscope. The characteristics of the oscilloscope are described as follows: bearing bandwidth of 60 MHz, 100 MHz, and 200 MHz; with sampling rate of 2 GS/s; having four 2 or 4 Channel acquisition modes; being applicable to

peak detection (12 ns Burr), sampling, averaging and single passing; being capable of setting menu and selecting waveforms automatically; owning waveform and setting memory (two or four 2500 point reference waveform, 10 front panel setting options); being provided with fast Fourier transform (FFT) function. When cable suffers external distribution, the oscilloscope displays the waveform. Basing the memorizing function of the oscilloscope, the waveforms stored can be used for analysis.
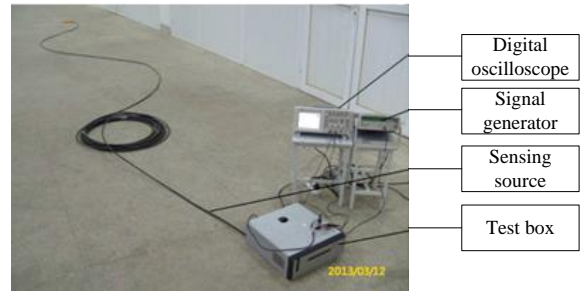
FIGURE 5 The test apparatus

To test the detecting ability of the system to the disturbance signals, the FFT transformation function of the digital oscilloscope was utilized in the test to convert the waveform of the signals into frequency. The frequency of the function signal generator was set at 20 Hz around. Figure 6 shows the frequency spectrum of the system in case of no external disturbance; the frequency spectrum of the interference wave tends to be monodromy and constant beat frequency.
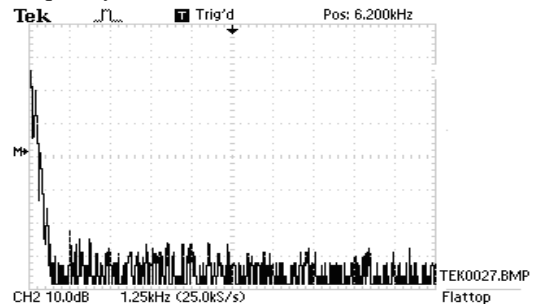
FIGURE 6 The frequency spectrum without continuous perturbation

By randomly applying discontinuous equal-amplitude disturbances on three places of the cable respectively, it is observable that the spectrums displayed by the oscilloscope exhibits significant variations in similar amplitude, as shown in Figure 7.

a)   b)   c)

FIGURE 7 The diagram of three discontinuous disturbance frequency spectrum

Figure 8 shows the frequency spectrum displayed by the oscilloscope after artificially applying continuous disturbances on other three places of the cable respectively. As it shown, under continuous vibration, part of the frequency spectrums shows variations in amplitude. The amplitude variations of the waves on the three positions are similar.
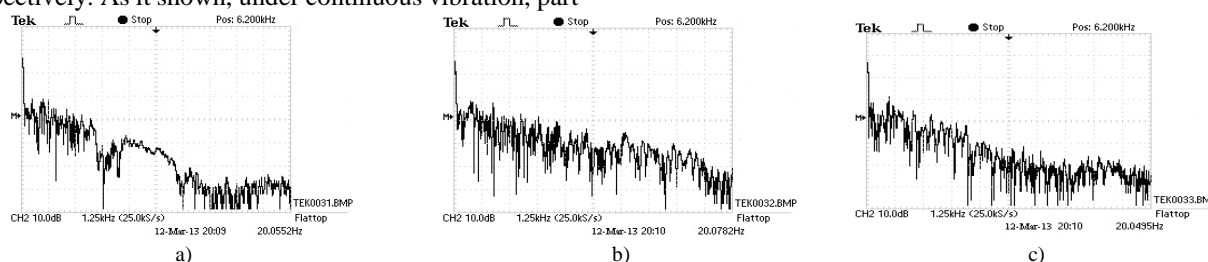
FIGURE 8 The frequency spectrum of three places at continuous perturbation

By comparing the frequency spectrums in Figures 7 and 8, it is easy to judge whether or not there are external disturbances.

5 Conclusions

This study proposed an M-Z interferometer -based disturbance detection interferometric fibre optic sensor system and analysed the sensing principle of the system. In this system, two single-mode optical fibres passing through the central tube of cable were used as the signal arm and reference arm of the M-Z interferometer. Subsequently, basing on corresponding working circuit, this study developed specific test device. Finally, the system was applied with disturbance test. The test result suggested that this system was very sensitive. Once there were minor disturbance on the cable, could the wave generate vibrations. Thus, it is easy to judge whether or not there are external disturbances using this system.

Acknowledgments

References

[1] Qiao Z, Zhao D, Jiang M, Hua N, Gu Z 2006 Stability analysis for final optical system of the additional beam of "Shengguang-II" laser facility *High Power Laser and Particle Beams* **18**(3) 419-22

[2] Trummer D J, Foley R J, Shaw G S 1998 Stability of optical element in NIF target area building UCRL-JC-129682

[3] Pechstedt R D and Jackson D A 1995 Design of acompliant-cylinder-type fibre-optic accelerometer: theory and experiment *Applied Optics* **34**(16) 3009-17

[4] Rivera H L, Garcia-Souto J A, Sanz J 2000 Measurements of mechanical vibrations at magnetic cores of power transformers with fibre-optic interferometric intrinsic sensor *IEEE Journal of Selected Topics in Quantum Electronics* 6(5) 788-97

[5] Garcia-Souto J A, and Rivera H L 2000 Multichannel fibre-optic interferometric sensor formeasurements of temperature and vibrations in composite materials *IEEE Journal on Selected Topics in Quantum Electronics* **6**(5) 780-7

[6] Pannell C N, Jones J D C, Jackson D A 1994 The effect of environmental acoustic noise on optical fibre based velocity and vibration sensor systems *Measurement Science & Technology* 5(4) 412-7

[7] Hu A, Rao Y, Nie Z, Niu Y 2004 Characteristics of the etched long-period fibre gratings *Acta Photonica Sinica* **33**(8) 916-9 *(in Chinese)*

[8] Liang Y et a1 2007 Optical fibre Sagnac interferometeric sensor for measurement of feeble vibration *Journal of Harbin Engineering University* **28**(1) 118-22 *(in Chinese)*

[9] Huang C, Cai H, Geng J et a1 2005 Wavelength interrogation based on a Math-Zehnder interferometer with a 3×3 fibre coupler for fibre Bragg grating sensors *Chinese Journal of Lasers* **32**(10) 1397-1400 *(in Chinese)*

[10] Li S, Zhu Y 2009 Fibre border guard system based the mach-zehnder optical fibre interferometer *Industrial Control Computer* 22(8) 36-8 *(in Chinese)*

[11] Wang D, Yang B et al 2006 The application of optical fibre sensors in the submarine cable monitoring, in *the First National Conference on Submarine Cable Communication Technology in Wuhan*, 2006

Authors

Youneng Zhang, born in November, 1973, Tongling, Anhui, China

Current position, grades: associate professor at Anhui Industry & Trade Vocational Technical College, Huainan, Anhui, China.
University studies: Master degree in computer application technology, Anhui University of Science and Technology, Huainan, Anhui, 2009.
Scientific interests: mine communication, RFID technology, internet of things technology, intelligent monitoring technology, and optical communication technology.

Dawen Wang, born in November, 1962, Huainan, Anhui, China

Current position, grades: the Experiment Center of Anhui Industry &Trade Vocational Technical College, Huainan, Anhui, China, Director of experiment center, Senior engineer.
University studies: Bachelor's degree department of electronic engineering of Nanjing institute of technology, 1986.
Scientific interests: optoelectronics and optical fibre sensing.

Li Ding, born in October, 1983, Lianshui, Jiangsu, China

Current position, grades: Anhui Industry & Trade Vocational Technical College, Huainan, Anhui, lecturer.
University studies: Master degree in control theory and engineering, Anhui University of Science and Technology, Huainan, Anhui, 2009.
Scientific interests: mine communication, electronic information, DSP Technology, applied electronics, electrical engineering, single chip microcomputer.
Experience: research on coal mine communication technology based on DSP, Anhui University of Science and Technology, 2009.

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

Numerical simulation of aerodynamic characteristics of windmill based on computer graphics display technology

Jingbin Hao*, Haifeng Yang, Zhongkai Li

College of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, China 221116

Abstract

This paper proposed research on wind computation fluid dynamics(CFD) analysis and visualization in scientific computing facing to OSG in order to obtain numerical result of aerodynamic characteristics stimulated by windmill and to display vivid and direct 3-D effect for easy observation experiment. This paper applied CFX module in ANSYS Workbench software to conduct CFD analysis and calculation of windmill through computer graphic display technology based on CFD principle and realize numerical stimulation of aerodynamic characteristics of windmill. Then the research on visualization in scientific computing of aerodynamic characteristics of windmill was realized combining with OSG software that face to virtual display technology

Keywords: computation fluid dynamics, computer graphics display technology, numerical stimulation, 3-D geometric model

1 Introduction

Various countries in the world are searching for new energy shortcut for the shortage of world energy [1]. Therefore, wind energy becomes an area, to which many countries pay high attention. For large volume and super high large wind driven generator, wind load is a quite important control load. However, problems such as wind pressure distribution and smooth interference usually cannot solved by normative solution pattern when important parameter of wind load is mentioned. Traditional action is to put entity or integration model into wind tunnel to conduct analysis and stimulation experiment of relative data so that obtain relative wind load experimental data result [2], which is used for design and improvement data of relative entity model. For instance, when airplane and car is done improved design, we usually extract relative experimental data in wind tunnel experiment [3]. However, wind tunnel has drawbacks of insufficient scale effect, high cost, long period and bad flexibility.

Scientific visualization has become the important research direction of current computation graphics [4]. People use traditional original visualization technology method to count or draw chart on scientific analysis data for observation experiment [5]. We aims to display 3-D flow field result figure generated by numerical simulation of aerodynamics characteristics of windmill in vivid and direct effect, make users observe the process when wind blowing over windmill more directly, increase perceptibility of windmill flow field and improve scientific research level of windmill. Therefore, we introduce "numerical wind tunnel" by combination of software and hardware for conducting numerical

simulation analysis technology of aerodynamics characteristics of relative model [6]. Car virtual wind tunnel research of Jili University and virtual wind tunnel research on aircraft of Beijing University of Aeronautics and Astronautics [7] have problem of few application cases and few departments that can provide 3-D data. Therefore, this paper explored and studied how to realize application of visualization in scientific computing more deeply.

2 Overview of computation fluid dynamics

Computation fluid dynamics (CFD) refers to systematic analysis and research on physical phenomenon such as heat conduction and fluid flow by computer numerical analysis and calculation and graphics display technology [8]. CFD is to conduct numerical simulation on fluid flow under the control of three flow basic equation [9]. Fundamental physical quantity in various position within fluid field and their basic situation of changing over time can be derived from these numerical simulation. CFD can display simulation result on computer screen by specific calculation means. In addition, it can also observe detailed effect in flow field to display flow scene directly and vividly.

Procedures of numerical simulation on fluid flow characteristics by CFD are as follows. First, a mathematical model that reflects problem nature should be set up. That is, to build up integro-differential equation that reflects the relationship between quantities of problem with definite condition and three conservation equation with their definite condition which is easy for calculation simulation solution. Then numerical discrete optimization method should be constructed as the core of

---

* *Corresponding author* e-mail: jingbinhjb@163.com

CFD, which is easy for research and analysis on CFD. Moreover, compilation and calculation are carried out. It includes division of computer network, import of initial conditions, boundary conditions, setting of control parameter, etc. At last, the result is conducted refine processing to obtain a high quality of display result.

The practical solution process of CFD is very complicated. The application of CFD is also complex. Therefore, many users select CFD commercial software. The structure of this commercial CFD software mainly includes pre-processing, solution and post processing.

3 Pre-processing and data conversion of 3-d geometric model of windmill

3.1 PREPROCESSING OF 3-D GEOMETRIC MODEL OF WINDMILL

Wind turbine blade and overall stand is selected to analyse and calculate and conduct numerical simulation research of aerodynamic characteristics of windmill.

3.1.1 Construction of Windmill 3-D Model

Windmill whose rated power is 1.5 megawatt designed by some enterprise was selected. And UG software produced by Siemens PLM Software Corporation was used for windmill 3-D geometric modelling. Fig.1 is UG windmill 3-D geometric model. Tower in tapered tubular steel structure was made of low-alloy high tensile structural steel Q 345. Hub centre is of 70 m high, 11 m/s rated wind speed, 3.5 m/s cut-in wind speed, 25 m/s cut-out wind speed and 3847 m2 swept area. CFX application module in ANSYS-Workbench was selected to do one way fluid-structure interaction analysis to realize calculation process of numerical simulation of aerodynamic characteristics of windmill because UG software itself is hard to realize that process.

3.1.2 Realization of UG and ANSYS Software Data Interface

UG is a software system of CAD and CAM that can realize interaction function. After 3-D model design software UG was used for constructing windmill 3-D geometric model, is was done data conversion and then applied into ANSYS-Workbench software to realize numerical simulation of aerodynamic characteristics of windmill. And IGES file format which is the sharing common format of UG and ANSYS-Workbench. Their characteristic was applied to read complete data from different platform to realize standard of data interchange between CAD and CAM system [10]. Thus, UG software could be used for deriving 3-D geometric model of windmill. Then it was derived into ANSYS-Workbench to conduct windmill 3-D geometric reconstruction to achieve numerical simulation of aerodynamic characteristics of windmill.

FIGURE 1 3-D geometric model of UG windmill

3.2 DATA CONVERSION OF WINDMILL 3-D GEOMETRIC MODEL FROM UG TO ANSYS

We have pre-processed windmill geometric model and realize interface function of data conversion of UG and ANSYS. Therefore, we can directly do data conversion on windmill 3-D geometric model into IGES file format and then import it into ANSYS Workbench software to do research on numerical simulation analysis of aerodynamic characteristics of windmill.

IGES file derived from windmill 3-D geometric model should be prepared before analysis. We selected Start Menu, through the path "ALL Programe, ANSYS14.0 > Workbench", start ANSYS Workbench module, open "fengjizhengti, igs" file in window of "File > Import" path and select Geometry to conduct "Update" geometric model load.

4 Application of CFD module from ANSYS software in numerical simulation of aerodynamic characteristics

Preparation of geometric file was basically fulfilled through the previous work. Next, CFX module in ANSYS software was applied to do numerical simulation of aerodynamic characteristics of windmill. CFD commercial software, which has regularity and systematicness was needed to solve the problem because CFD calculation involves allocation of computer software and hardware and the complexity of problem. CFX was constantly integrated the most advanced global physical geometrical model and simulation solution algorithm for ensuring the accuracy and stability of calculation result. Meanwhile, finite volume method based on finite element was also used for ensuring basic conservation characteristics and accuracy of numerical value [11].

4.1 REALIZATION OF PREPROCESSING OF WINDMILL 3-D GEOMETRIC MODEL IN CFX MODULE OF ANSYS SOFTWARE

Pre-processing of 3-D geometric model mainly involves functional operation in DesignModeler on the model that has been imported to realize modelling process of rotating field and external flow field.

FIGURE 2 windmill model of cylindrical rotating field
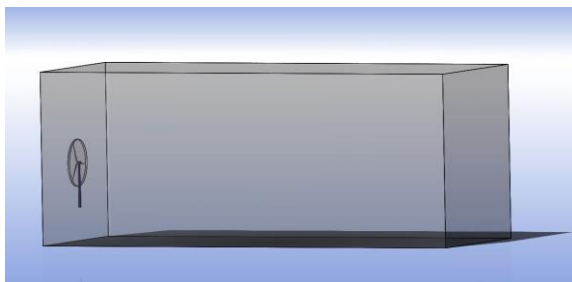
FIGURE 3 square external flow field of windmill
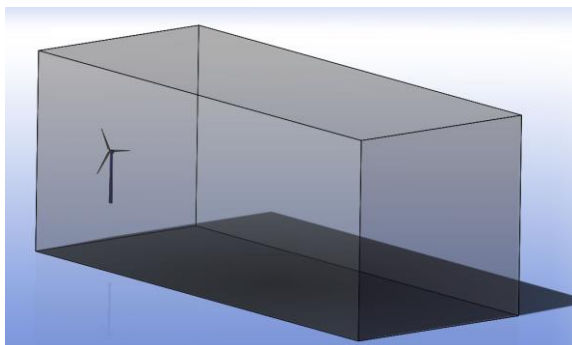
FIGURE 4 square external flow field of windmill after Boolean calculation

First, we double clicked GeometryA2 module to conduct windmill 3-D geometric model load to enter DesignModeler analysis module and selected Meter unit system in "unit" dialog box. Second, we constructed cylindrical rotating field of windmill 3-D geometric model through path "Create > Primitives > Cylinder". Fig.2 is windmill model of cylindrical rotating field. Then

blade model was removed. We did Boolean calculation of windmill 3-D geometric model through path "Create > Primitives > Box". We selected "Yes" order in Preserve Tool Bodies for reservation of blade model. At last, after rotating field setting, square external flow field was constructed through path "Create > Primitives > Box". Then Boolean calculation was done again. Fig.3 is square external flow field. Fig.4 is square external field after Boolean calculation.

4.2 FLOW FIELD MESH DIVISION OF 3-D GEOMETRIC MODEL OF WINDMILL

After pre-processing of 3-D geometric model of windmill by ANSYS Workbench, we used Mesh module to divide 3-D geometric model mesh of windmill to generate the needed calculation mesh.

First, we constructed Mesh module, dragged GeometryA2 into MeshB2 and imported pre-processed windmill 3-D geometric model into Mesh module. Then we expanded path "Project > Model > Geometry" to observe some Solid body. Windmill model was suppressed and only the flow field was remained. Solid body would show out X mark after suppression. Then we revised Solver Preference value as CFX through path "Project > Model (B3) > Mesh". CFX-meshing under Mesh was right clicked. At last, we realized the setting of boundary condition by setting entrance of external flow field, exit of reserve side of windmill, ground and part surface of windmill stand. Meanwhile, some specific network properties were set. At last, mesh model of windmill flow field is generated as shown in Fig. 5.
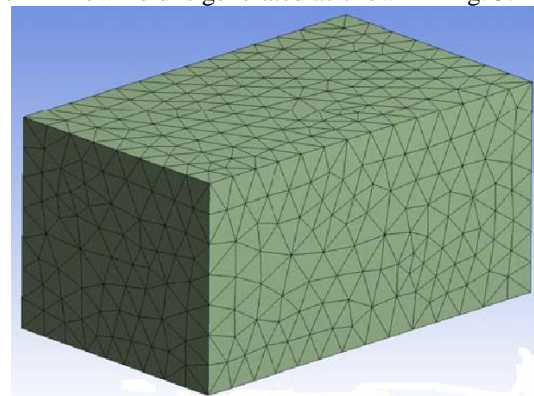
FIGURE 5 Mesh model of windmill flow field
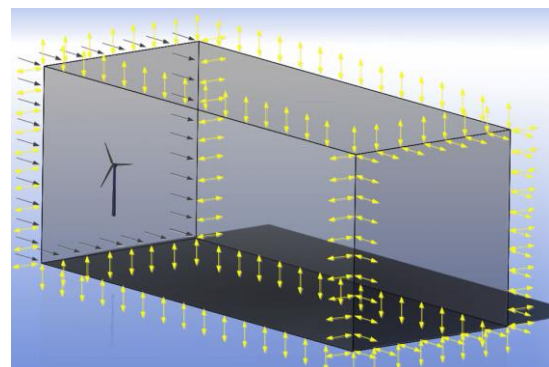
FIGURE 6 load figure of rotating field and external flow field

4.3 SETTING AND SOLUTION OF FLUID ANALYSIS OF WINDMILL 3-D GEOMETRIC MODEL

After division of windmill 3-D geometric model mesh, we entered into fluid analysis setting including designing boundary condition and properties of solver.

First, we set analysis category and imported divided Mesh file into CFX-Pre for fluid calculation. Then CFX module was generated automatically and connected. Then rotating field and external fluid field were constructed. At last, the relative boundary parameter condition was set. Fig. 6 is the load figure of rotating field and external fluid flow field after setting of relative parameters.

FIGURE 7 windmill 3-D flow field

FIGURE 8 flow line of 3-D flow field removing windmill

We only need to do simple solver setting after calculation of fluid and then solution result could be achieved. Fig. 7 is windmill 3-D flow line after setting of solver.

Through application of CFX module in ANSYS Workbench software, 3-D flow field figure removing windmill can also be independently produced. Fig. 8 is flow line of 3-D flow field removing windmill.

5 Conclusion

This paper mainly illustrated application and basic knowledge of common commercial software in CFD in order to study windmill CFD. Based on it, we pre-processed windmill geometric model and realized interface function of data conversion between UG and ANSYS. Windmill 3-D geometric model constructed in UG was directly conversed into IGES format and then imported into ANSYSWorkbench software for study of numerical simulation of aerodynamic characteristics of windmill. IGES file derived from windmill 3-D geometric model should be prepared before analysis. And CFD module in ANSYS software was applied for numerical simulation of aerodynamic characteristics of windmill. CFX module can ensure accuracy and stability of calculation result as well as basic conservation characteristics and accuracy of numerical value. Pre-processing of 3-D geometric model mainly involved functional operation in DesignModeler on model that have been imported to realize the modelling process of 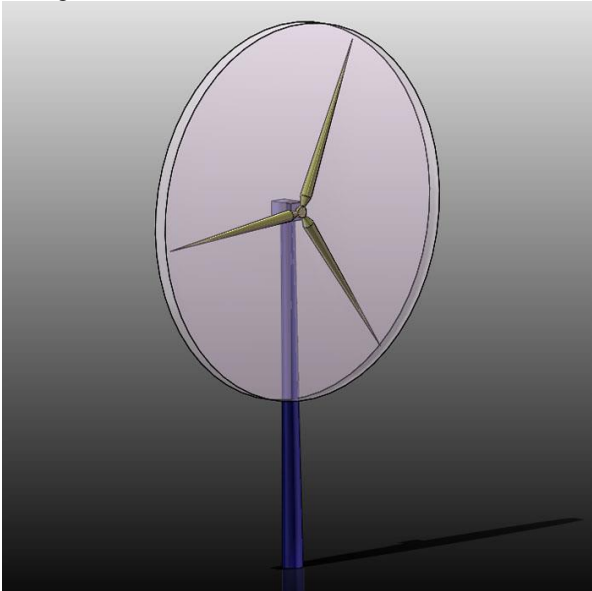rotating field and external flow field. Through pre-processing of windmill 3-D geometric model by ANSYS Workbench, we used Mesh module to divide the mesh of windmill 3-D geometric model to generate the needed calculation mesh. After division of windmill 3-D geometric model mesh, we entered into fluid analysis setting including designing boundary condition and properties of solver. We only need to do simple solver setting after calculation of fluid and then solution result could be achieved. At last, the numerical simulation of aerodynamic characteristics of windmill was achieved.

Acknowledgment

References

[1] Fengyi C, Shaohua G 2012 Draught fan industry market analysis and prediction *General Machinery* **2** 22-26
[2] Tianshu Z 2007 Research on numerical wind tunnel method of shape factor confirmation of wind load. *Tianjin University*
[3] Xiaobi S 2010 Aerodynamic computation on wind pressure distribution of ultra-large cooling tower Guangdong Electronic Design Institute *Journal of Wuhan University* **8** 270-274

[4] Huiyan J, Ning L 2007 3-D reconstructed system design based on advanced ray casting method Collected Papers of 2nd 3-D Image Technology and Application (International) Seminar **9** 18-20
[5] Huisen Z 2006 3-D visualization platform design and realization facing to tourist resource management *Shanxi Normal University* **5** 3-4
[6] Tianqin G 2009 Low speed wind tunnel boundary layer control numerical simulation research *Huazhong University of Science and Technology* **5** 1-3

[7] Tiezhong L, Qiang L, Zhoufu L 2009 Virtual aircraft wind tunnel research and realization Beijing University of Aeronautics and Astronautics. *Journal of System Simulation* **14** 1-3

[8] Jian L 2002 3-D visualization processing of engine crankshaft element calculation based on OPENGL technology *Shanghai Jiaotong University* **2** 1-10

[9] Tuncer C, Jian P S, Fassi K, Eric L 2009 Computational fluid dynamics for engineers *Beijing: Tsinghua University Press* **4** 31-35

[10] Jining S 2010 Basic application course of ANSYS CFX convection heat transfer numerical simulation. *National Defence Industry Press*.

[11] He Y, Xiang J 2003 Computer graphics principle, method and application. *Higher Education Press*.

Authors

Jingbin Hao, born in June, 1982, Jiangsu Province of China

Current position, grades: lecturer
University studies: Ph. D degree was earned in major of Mechanical Manufacture and Automation, University of Mining and Technology in 2011.
Scientific interest: laser rapid prototyping, machine vision and laser imaging.

Haifeng Yang, born in May, 1981, Jiangsu Province of Chin

Current position, grades: Associate Professor
University studies: Ph. D degree was earned in major of Materials Science, Jiangsu University in 2009.
Scientific interest: Laser nano-manufacturing, Unconventional Machining Technology.

Zhongkai Li, born in February, 1980, Shandong Province of China

Current position, grades: Associate Professor
University studies: Ph. D degree was earned in major of Mechanical Design and Theory, Zhejiang University in 2009.
Scientific interest: Product family design, Enterprise informatization

Authors' index

Name	Page	Name	Page	Name	Page	Name	Page
Baranova S F	435	Jin Bo	450	Ning Xiaobin	528	Xu Jianmin	34
Bin Cheng	100	Kou Guang Xiao	438	Pan Peng	186	Xu Wensu	236
Cai Hui	145	Lei Ming	19	Pan Weili	95	Xu Zi-heng	423
Cai Xiaojun	186	Li Hanshan	19	Peng Haican	52	Xue Yanru	312
Cao Jing	312	Li Huaping	177	Peng Sui	423	Xue Yun-tao	423
Cao Qiao Ying	438	Li Lin	284	Pravednaya N P	435	Yan Wei	365
Cao Wangcheng	236	Li Ning	528	Qi Yan	153	Yan Xiaoheng	491
Che Xiaoyi	77	Li Ping	543	Qian Huanyan	12	Yang Bin	256
Chen Caixian	371	Li Shenglin	145	Qian Limiao	65	Yang Bingru	277
Chen Fei	358	Li Wei	196	Qing Xuhua	518	Yang Feng	303
Chen Guifang	181	Li Xiaoyu	116	Qiu Jianping	181	Yang Haifeng	560
Chen Jian	58	Li Xingchen	145	Ren Lei	240	Yang Hua-min	153
Chen Jiguang	12	Li Xinping	90	Ren Xuping	131	Yang Jing	167
Chen Lichao	181	Li Xiuli	339	Shan Yazhou	462	Yang Juan	26
Chen Long	358	Li Yi	358	Shi Jian-gang	404	Yang Liming	249
Chen Mingming	19	Li Yongchi	543	Shu Zouqi	503	Yang Qin	380
Chen Weihua	491	Li Yuanchun	468	Song Qiang	272	Yang Zhenyu	350
Chen Zhiding	218	Li Yunfa	131	Song Yifan	514	Yao Kun	52
Cheng Weijun	124	Li Zhongkai	560	Sun Xiaolei	203	Yao Yinghua	312
Cong Cong	518	Liang Chunhun	468	Tan Jie	265	Ying Yating	160
Cui Jingfei	214	Liao Zilong	514	Tang Song	534	You Xindong	131
Dai Linlin	410	Lin Junting	415	Tang Wei	90	Yu Xian	58
Dang Jianwu	415	Lin Peng-Chun	191	Tang Zhihong	58	Yu Xin	46
Deng Lixia	550	Liu Fangai	303	Tao Yangyang	497	Yu Yun-jun	423
Ding Li	555	Liu Hongtao	71	Tong Chao	423	Yuan Wei	103
Dong Kuiling	236	Liu Hongzhao	455	Wang Ai-min	272	Zhai Yun	277
Dong Suyu	167	Liu Jianying	550	Wang Chengduan	290	Zhan Jianjun	365
Dou Shasha	82	Liu Jin-can	404	Wang Chunhui	522	Zhang Chunfu	534
Du Shuang	484	Liu Kangsheng	46	Wang Dai	116	Zhang Haiping	131
Duan Shiwei	543	Liu Kun	228	Wang Dawen	555	Zhang Heng	145
Fan Qiang	223	Liu Lizheng	303	Wang Fei	40	Zhang Hongyang	477
Feng Qiaojuan	345	Liu Min	312	Wang Han Qing	438	Zhang Ji	207
Feng Wuwei	538	Liu Mingtao	543	Wang Jian	139	Zhang Jianqi	103
Fu Hua	491	Liu Ning	100	Wang Jianguo	518	Zhang Min	223
Gao Xiuju	249	Liu Qian	218	Wang Jinlin	214	Zhang Qiaoxin	550
Gu Lizhi	34	Liu Qiyuan	77	Wang Le	139	Zhang Youneng	555
Gu Wei Li	438	Liu Wei	71	Wang lei	510	Zhang Yue	203
Guo Zhongxiao	514	Liu Wei	455	Wang Leilei	236	Zhang Yuhong	339
Han Fenglei	522	Liu Yachong	522	Wang Peng	153	Zhang Zhen	214
Han Huijian	371	Liu Ying	46	Wang Shunye	319	Zhao Lijun	450
Han Xiao	518	Liu Yutian	109	Wang Xiaochun	58	Zhao Sujuan	339
Han Xuesong	26	Liu Zheng	371	Wang Xiaoming	415	Zhao Xiaochun	7
Hao Jingbin	560	Long Kai	65	Wang Yanwen	249	Zhao Xuezuo	312
Hartnett Michael	240	Long Yinhui	514	Wang Yongjiao	345	Zhao Yan	167
He Chunyu	445	Lu Guancheng	58	Wang Yuansheng	228	Zhao Yang	265
He Dongjian	223	Lu Hsin-Ke	191	Wang Zhiming	334	Zheng Jian-Feng	207
He Feng	52	Lu Huaiwei	462	Wang Zhongshuang	497, 503	Zheng Lin	236
He Liping	65	Lu Huanda	46	Wang Zijing	65	Zheng Yong-jun	40
Hu Ankang	522	Luo Lei	228	Wei Hongxia	298	Zheng Yulan	329
Hu Junjie	109	Luo Shanming	34	Wu Di	445	Zhou Changyao	380
Hu Lianhua	90	Luo Tianqi	462	Wu Huarui	228	Zhou Feng	160
Hu Zhigang	71	Luo Youxin	77	Wu Kaijun	462	Zhou Jian	203
Huang Ning	203	Ma Nan	277	Wu Kehe	358	Zhu Shan-an	40
Huang Ruiyuan	543	Ma Weihong	228	Wu Tianyan	365	Zhu Shiqiang	450
Huang You-fang	256	Meng Yanmei	58	Wu Xiaoli	160	Zhu Shuzhen	334
Jiang Junping	528	Miao Jianguo	534	Xia Jiansheng	82	Zou Qilun	218
Jiao Zhijie	445	Nash Stephen	240	Xu Dong-min	256	Zuo Chuncheng	484

Cumulative Index

Mathematical and Computer Modelling

Xiaochun Zhao Distributed systems software architecture modelling and research methods
*Computer Modelling & New Technologies 2014 **18**(9) 7-11*

With the development of computer network technology, the open, heterogeneous and distributed systems have become the mainstream in current computer applications because of the sharing resources, high availability, parallel processing and so on. However, due to the problems of development, which are constant expansion of systems size, evolution and continuous improvement, maintenance that required, specific distribution, autonomy and heterogeneity, a lot of research and software development practice shows, the introduction of software architecture which guide distributed system to develop and assume component blueprint is a practical and effective way to solve the difficulties of the development of distributed systems and build distributed systems successfully. Therefore, how to improve the quality and efficiency of distributed systems development by using software architecture, and ensuring system maintenance and space evolution are the key to develop distributed systems, also the core of this study. Software architecture, formal description of distributed systems interaction style, refinement and mapping architecture, distributed architecture systems development methods under evolution and reconstruction driving were studied based on the current distributed systems development methods as well as the problem of inadequate means.

Keywords: Software architecture, distributed systems, interaction style, software architecture refinement

Jiguang Chen, Huanyan Qian Three-dimensional deployment algorithm based on ideal fluid dynamics model for mobile sensor networks
*Computer Modelling & New Technologies 2014 **18**(9) 12-18*

In the paper, a three-dimensional deployment algorithm based on ideal fluid model for sensor networks is proposed. On this basis, the proposed ideal fluid model is analysed, and the concept of flow field model is applied in deployment of wireless sensor networks. Sensor networks are abstracted as ideal fluid, with nodes as fluid micelles. In the deployment process, motion of nodes follows momentum conservation law of fluid micelle. Moreover, a simulation experiment is performed in this paper with the proposed deployment algorithm as the experimental subject. Coverage and uniformity are 2 indexes employed to evaluate the performance of the proposed algorithm. Shown by the simulation result, the three-dimensional deployment algorithm based on ideal fluid model for sensor networks leads to good deployment effect.

Keywords: three-dimensional deployment, mobile sensor networks, ideal fluid model

Ming Lei, Mingming Chen, Hanshan Li Goal's three-dimensional trajectory reconstruction based on the adaptive multiple target surface iteration method
*Computer Modelling & New Technologies 2014 **18**(9) 19-25*

In order to solving the problem of the multiple bullet target matching and multiple bullet three-dimensional trajectory calculation., this article proposes a method for bullet three-dimensional trajectory measurement based on an adaptive multiple target surface and multiple iteration algorithm. In this method, it set up a virtual cuboids target space which includes multiple bullet trajectories. It used two high speed cameras, one in left and another in right, to capture images. In the vertical direction of the bullet flying, these images were segmented into infinite plane, which are multiple target surfaces. It used the projective transformation to recover the target two-dimensional image from the left and the right images, and repeated this process until finding the bullet point in left view is coincide with that in right view. If the points were found, it indicated that it achieves the goal's accurately matching and every space location of the bullet. The simulation experimental results show that this method is feasible, image processing and analysis is merely influenced by background. This method can effectively realize the bullet three-dimensional trajectory target matching, and it has strong maneuverability.

Keywords: three-dimensional trajectory, target matching, perspective transformation, multiple iterations, adaptive multiple target surface

Juan Yang, Xuesong Han Verification of calculation efficiency of a new CS-PSO algorithm and its

application
*Computer Modelling & New Technologies 2014 **18**(9) 26-33*

Intelligent algorithm is developing rapidly with the development of computer technology. And it is widely used in scientific research and industrial application. As a kind of intelligent algorithm, particle swarm optimization (PSO) has been used in solving problem for a long time. It is based on the bird group behaviour and uses biological group model to find the optimal solution. Its advantages are fast calculation speed and easy implementation while the disadvantages are easily getting into the local extreme, slow convergence speed in the late evolutionary and poor precision. In order to avoid the disadvantages, some modification has been studied for PSO algorithm and establishes the concentration degree and steady degree based PSO (CS-PSO) algorithm in the paper. Based on the convergence performance of particle swarm depends on the particle exploration ability, search space has been adaptively adjusted to improve the convergence performance of particle swarm optimization with the variation of optimal fitness value. Corresponding adjusted method has been shown in the paper. According to the example verification, the CS-PSO is effective and then the algorithm is used in the bellow structure optimization.

Keywords: intelligent algorithm, particle swarm optimization, space adjusted, experimental test

Jianmin Xu, Lizhi Gu, Shanming Luo Structure optimization of cycloid gear based on the finite element method
*Computer Modelling & New Technologies 2014 **18**(9) 34-39*

A numerical model of cycloidal gear is created by using three-dimensional software and finite element analysis is applied with ANSYS platform. The first six natural frequencies and mode shapes are obtained, as a result. Influences from structure, material and thickness of the gear are investigated. Analysis shows that, modal shapes of cycloid gear are mainly circumferential modes, umbrella-type modes, torsional vibration mode and radial modes. The first six natural frequencies of 5 kinds of cycloid gear with variable cross-section were smaller than those of ordinary cycloid gears, and cycloid gears with variable cross-section can avoid resonance frequencies easily. Dynamics of five new cycloidal gears with variable cross-sections are consistent with ordinary cycloid gears. Modal frequencies of ordinary cycloid gears increases in accordance with materials, such as bearing steel, alloy steel and plastics; also, natural frequency increases with the increase of the thickness of the gear. Conclusions of this paper provide a basis for dynamic designing of cycloid gears.

Keywords: cycloid gear, finite element modal analysis, free modal, constraint modal

Yong-jun Zheng, Fei Wang, Shan-an Zhu Stochastic resonance induced by over-damped fractional Langevin equation with α-stable noise
*Computer Modelling & New Technologies 2014 **18**(9) 40-45*

Stochastic resonance phenomenon induced in a system described by over-damped fractional Langevin equation with α-stable noise is investigated. When there is no external α-stable noise, the stochastic resonance is observed in case of the fractional order less than one certain threshold. By applying α-stable noise, the influences of the noise intensity and characteristic exponent of α-stable noise on the occurrence of stochastic resonance phenomenon are characterized. We find that the proper noise intensity enlarges the peak value of output power spectrum which is significant for stochastic resonance. Adjusting the noise intensity, the behaviour of signal-to-noise ratio is non-monotonic and with a maximum value. Under the same conditions, the lower value of characteristic exponent of α-stable noise leads to the smaller noise intensity to achieve stochastic resonance.

Keywords: stochastic resonance, over-damped fractional Langevin equation, α-stable noise

Huanda Lu, Xin Yu, Ying Liu, Kangsheng Liu The simulation and analysis of fish school behaviours with different body lengths
*Computer Modelling & New Technologies 2014 **18**(9) 46-51*

In this paper, through simulating and modelling of fish school with different body length, we study the influences of the body length difference of fish school on spatial structure and group behaviour. Based on attraction/repulsion model, we obtain three typical spatial structure of group with different model parameters: mixture structure, periphery

structure and front-back structure. Moreover, we analyse the polarization index and average angular speed of group with different model parameters and get the corresponding relationship between these indices and model parameters. The results obtained in this paper coincide with the phenomenon observed in the natural world and the methods provide an effective way to study the fish school behaviour.

Keywords: computer simulation, attraction/repulsion model, aggregation behaviour, spatial structure

Feng He, Haican Peng, Kun Yao Research on the software trustworthiness extended measurement based on IMC
*Computer Modelling & New Technologies 2014 **18**(9) 52-57*

This paper introduces two methods of extension measure based on model checking algorithm of interactive Markov chains (IMC) to decide the software trustworthiness. The first extended measurement is to establish multiple corresponding temporal logic relations for each software trustworthy attribute that affecting software trustworthiness, also is to use multiple temporal logic to describe a software trustworthy attribute, which is aim to measure the software trustworthiness on the multi-level and fine-grained. Then the paper will determine the measurement ultimately. The second extended measurement is to locate for the untrusted states, then find out the detail path and detail parameters of the path. Next, we will get the location that not trusted through further analysis. Eventually meet people's expectations by improving.

Keywords: software trustworthiness, model checking, finite state machine model, trustworthy attribute

Guancheng Lu, Yanmei Meng, Jian Chen, Zhihong Tang, Xiaochun Wang, Xian Yu A nonlinear system modelling approach to industrial cane sugar crystallization
*Computer Modelling & New Technologies 2014 **18**(9) 58-64*

Cane sugar crystallization is a non-linear process where multiple control parameters are involved, which makes it rather difficult to reveal its internal mechanism by mechanism modelling. Derived from variants of standard support vector machine method, an online control system modelling method based on multi-input and multi-output proximal least square support vector machine is proposed to be applied in sugar crystallization process. This method takes multiple process control parameters as the input and output of machine learning algorithm, through which the inherent law between key and auxiliary parameters in the sugar crystallization process is established. The ultimate goal is to control the sugar crystallization process automatically. The experimental results show that the accuracy rate of the model output is 95%.

Keywords: multi-input and multi-output proximal least square SVM, sugar crystallization control system, machine learning, nonlinear modelling

Liping He, Limiao Qian, Zijing Wang, Kai Long Finite element modelling and modal analysis of structure with bolted joints
*Computer Modelling & New Technologies 2014 **18**(9) 65-70*

In order to investigate a modal analysis of structure with bolted joints, five kinds of finite element models, a solid bolt model, a no-bolt model, a modified bolt model, a beam bolt model and a rigid bolt model, are introduced. Among these models, the solid bolt model, modified bolt model and rigid bolt model provide good accurate responses compared with the modal test results. The solid bolt model, which can also be used for stress analysis, is recommended in view of applicability. The rigid bolt model is recommended in view of usefulness and effectiveness. For bolted flanged connection, the magnitude of bolt pretension has little influence to structural modes. To improve the calculation accuracy of higher modes, a better idea is to modify the finite model by modal test.

Keywords: bolted joints, flange, pretension, modal, finite element analysis

Wei Liu, Zhigang Hu, Hongtao Liu An automatic approach to detecting and eliminating lazy classes based on abstract syntax trees
*Computer Modelling & New Technologies 2014 **18**(9) 71-76*

To detect and eliminate lazy classes in source code, an automatic approach based on abstract syntax trees (ASTs) is proposed. Source code files transform to ASTs at first, then the relationships between classes are extracted from the

ASTs. Three common relationships are considered, which are generalization, association and dependency. Some definitions are proposed to represent the classes set of different kinds of relationships. After carrying out several set operations on these sets, the candidate lazy classes set is obtained. By further manual examination, the true lazy classes are acquired. Finally, a specific lazy class will be removed automatically from the project. Four projects are tested to detect and eliminate the lazy classes. The experimental results show that the proposed detection algorithm has high precision rate. In addition, this approach has good efficiency, and its execution time has a linear relationship to the size of a system.

Keywords: code smells, lazy classes, abstract syntax trees, refactoring, classes set

Youxin Luo, Qiyuan Liu, Xiaoyi Che New information multivariable optimization MGM(1,*n*) model with non equidistance and based on background value optimization
*Computer Modelling & New Technologies 2014 **18**(9) 77-81*

The function with non-homogeneous exponential law, based on index characteristic and integral characteristic of grey model GM(1,1), was used to fit the one-time accumulated sequence, and the formula of background value was given, aiming at the problem of lower precision as well as lower adaptability in non-equidistant multivariable model MGM(1,*n*). A new information optimization model MGM(1,*n*) with non equidistance and multi variable based on background optimization was put forward, took the *m-th* component of the original sequence as initial condition, the mean relative error as objective function, and the modified one of initial value and the parameters of background value as design variables. This proposed MGM(1,*n*) model can be used in equidistance & non-equidistance modelling with higher precision as well as stronger adaptability. Examples have validated the practicability and reliability.

Keywords: multivariable, background value; optimizing, new information, non-equidistance sequence, non-equidistant MGM(1,*n*) model, least square method

Jiansheng Xia, Shasha Dou The study on elliptical flange hole forming based on finite element analysis
*Computer Modelling & New Technologies **18**(9) 82-89*

The flange hole forming is a complex process, Under the assumption of Prandtl-Reuss flow rule and von Mises yield criterion, the incremental elasto-plastic large deformation finite element model was established based on the Updated Lagrangian Formulation (ULF).The elasto-plastic conversions of boundary and deformation are reduced with r-min rule. The friction phenomenon of slippage and viscosity at the boundary interface is revised with increment of revision Coulomb rule. The increment rules are led into the whole stiffness matrix, and derived out the stiffness equation. The studies show that the influence on steel elliptical hole flange forming deformation is influenced by punch structure and parameter. The dates show that finite element simulation and experimental result have a good consistency.

Keywords: elasto-plastic, FEM Simulation, elliptical hole flange

Lianhua Hu, Xinping Li, Wei Tang Adaptive IMC for variable parameter systems with large dead-times
*Computer Modelling & New Technologies 2014 **18**(9) 90-94*

Dead-time variation can cause object mismatches in traditional internal model control (IMC) systems and may result in a significant overshoot. The adjustment time may increase due to this variation and sometimes even causes oscillation instability. An adaptive IMC method is proposed in this paper to solve the problem of variable parameters in the control process. The adaptive law is designed to optimize local parameters relative to the output error of both the plant and model, ensuring that the model approximates the real plant. The control structure adopts IMC and simulation results show that this type of control structure exhibits some promising characteristics, such as high accuracy, robustness and disturbance rejection. This model is therefore suitable for systems with large dead-time varying parameters.

Keywords: Large dead-time, Internal model control (IMC), Adaptive, Variable parameter

Computer and Information Technologies

Weili Pan Research on construction of normative ideological instruction teaching management system based on ZigBee wireless sensor network
*Computer Modelling & New Technologies 2014 **18**(9) 95-99*

In the process of standard ideological and political teaching management, teaching quality management, teacher

management, student management and education environment management etc are all very important links. In order to establish and perfect the management system of ideological and political teaching in school, management utility should be developed. And the system optimizing should be comprehensively enhanced in teaching. We are to construct a management system on the basis of ZigBee through wireless sensor, RMON (remote monitoring) of ideological instruction teaching supervision, multimedia teaching and classroom environment management etc. will be designed through coordinating with Go-Ahead technology. The corresponding hardware and software facilities will be designed and developed. The role of wireless sensors in teaching management will be reflected through the image of flow image. Standardized teaching management would be genuinely realized with its advantages of convenience, utility, high efficiency, and energy conservation.

Keywords: teaching management, wireless sensor, ZigBee, RMON (remote monitoring)

Ning Liu, Cheng Bin Hotel room design based on face recognition, environment monitoring and AV regulating system
*Computer Modelling & New Technologies 2014 **18**(9) 100-102*

With the improvement of people's living standard, the tourism industry has transformed from the traditional sightseeing tourism to the vacation tourism, In order to improve the quality of the overall experience of resort hotel, the article by using the image acquisition technology, the facial expression recognition technology, the changes of Gabor and the DIB support, VC++ of bitmap provided by Visual Studio, has designed the basic member variables required by bitmap processing packaging DIB and the unrelated CDib of the member function device, and has constructed the facial expression recognition module of the mood regulation system of hotel interior. On the side of audio stream, use PC port as the audio stream scheme of the sound source, in the PC port, separate Av signal, preset the operation and treatment on the end of the playback software, and make self-regulation of sound adjustment in the output end of sound. The highly fidelity of sound has been established.

Keywords: hotel interior, Influence acquisition, feature extraction, frequency

Wei Yuan, Jianqi Zhang The study of eccentric bunghole self-positioning system based on computer vision technology
*Computer Modelling & New Technologies 2014 **18**(9) 103-108*

This paper on the basis of deeply understanding the domestic automatic filling equipment's, develops an eccentric bunghole self-positioning system based on computer vision technology, through full automatic camera calibration and improved Hoff conversion algorithm to obtain the coordinate information of circle centre of eccentric fillers, thus controls the operation of driving stepper motor, realizes the automatic positioning of sprue guns to eccentric bunghole. The experiment proves the operation accuracy of this system is high, the speed is fast, the algorithm is efficient, stable, and a high practical value.

Keywords: computer vision technology, eccentric bunghole, automatic positioning, camera calibration, Hoff transformation algorithm

Yutian Liu, Junjie Hu The analysis and avoidance of fault agent in flocking of multi-agent
*Computer Modelling & New Technologies 2014 **18**(9) 109-115*

In the last decade considerable research efforts have been spent to the motion of flocking of multi-agent. Special attention has been put in the applications, especially for those operations in real environment where a high degree of safety as well as self-diagnostics capabilities are required. The development of effective strategies of fault diagnosis for flocking of multi-agent is a critical research task. In this paper, the flocking motion of multi-agent with a leader is studied. When flocking in the real environment, it is inevitable for agents to occur faults. The faults occurred in different agent will lead to different effects for flocking. According to the variety of the velocity of agents, the fault types are classified. A fault agent avoidance method is proposed and implemented in a multi-agent flocking system. The simulation results show the method can help the agents to avoid the fault agent.

Keywords: multi-agent, flocking, fault agent, fault agent avoidance, leader

Xiaoyu Li, Dai Wang Quantum public-key cryptosystem without quantum channel between any two users

based on the Bell state measurement
*Computer Modelling & New Technologies 2014 **18**(9) 116-123*

In this paper, a quantum public-key cryptosystem without quantum channel between any two users based on the Bell state measurement is presented. A user Alice shares a set of Einstein-Podolsky-Rosen (EPR) pairs with key management centre (KMC) as the private key and the public key. By performing the Bell state measurement on the public key and the auxiliary qubits any other user can send encrypted message to Alice. On the other hand, digital signature can also be achieved by this public-key cryptosystem. The laws of quantum physics guarantee the unconditional security of this public-key cryptosystem. No quantum channels are needed between any two users. So it is easier to carry out in practice and more robust against possible attacks.

Keywords: public-key, quantum cryptography, EPR pair, the Bell state measurement, digital signature

Weijun Cheng Dual-hop variable gain relaying in mixed multipath/shadowing fading channels
*Computer Modelling & New Technologies **18**(9) 124-130*

In this paper, we investigated the end-to-end performance of a dual-hop variable gain relaying system over mixed fading environment. In such environment, the wireless links of relaying system undergo different fading conditions, where one link is subject to the Nakagami-m fading, the other link is subject to the composite Nakagami-lognormal fading which is approximated by using mixture gamma fading model. Based on the cumulative distribution function of the end-to-end signal-to-noise ratio (SNR), some novel closed-form expressions of the average end-to-end SNR, the outage probability, the symbol error rate and the ergodic capacity for the dual-hop variable gain system are derived, respectively. Then, some approximate analysis and the diversity order are found based on the above new expressions in high SNR region. Finally, numerical and simulation results are shown to verify the accuracy of the theoretical analysis.

Keywords: dual-hop relaying, mixed fading channels, mixture Gamma distribution, performance analysis

Xuping Ren, Haiping Zhang, Yunfa Li, Xindong You Scalable authentication protocol in RFID-based systems
*Computer Modelling & New Technologies 2014 **18**(9) 131-138*

Because of vast highly sensitive business information within RFID system, there is an urgent need for an effective and secure protocol to ensure the interests of various stakeholders. In this paper, we propose a scalable authentication protocol to provide classification protection. GNY logic formal approach is used to verify the design correctness of the protocol. The performance is evaluated and compared with other related protocols in three aspects: storage, computation requirement and communication overload. The analysis shows that the proposed protocol need less computation requirement and memory with acceptable communication overload. The conclusion indicates that the protocol is reliable and more scalable in RFID-based sensor systems.

Keywords: RFID security, authentication protocol, scalable

Jian Wang, Le Wang A new anonymity-based protocol preserving privacy based cloud environment
*Computer Modelling & New Technologies 2014 **18**(9) 139-144*

With the development of cloud computing application, more and more people would like to do business under this environment. But more attention should be paid to the disclosure of privacy during the transaction. Customers will reject to do business on the cloud platform if the cloud environment cannot avoid disclosing their private data. Nowadays, little work has been done about how to prevent sensitive attributes leaking between service providers. Therefore, this paper proposed a new anonymity-based protocol to protect privacy.

Keywords: anonymity, cloud computing, privacy preserving, sensitive attributes, protocol

Xingchen Li, Shenglin Li, Heng Zhang, Hui Cai The research of data conflict in digital camp management and control system based on IOT
*Computer Modelling & New Technologies 2014 **18**(9) 145-152*

A core technology in IOT (Internet of Thing) is data processing and there may be a mismatch between data collection and data interface. Aiming at data conflicts in the DCMCS (Digital Camp Management Control System), this paper defines data conflicts with formal description and divides the data conflicts of the DCMCS into six types. For each

type of the data conflicts, a resolution strategy is designed to solve it. Then the paper designs the DCRS (Data Conflict Resolution System) to detect and resolve the data conflicts automatically and quickly. The solutions provide technical support for data interaction, which have been successfully verified in the DCMCS.

Keywords: formal description, data conflict, resolution strategy, DCRS

Peng Wang, Yan Qi, Hua-min Yang Analysis and study on the performance of query based on NOSQL database
Computer Modelling & New Technologies 2014 **18**(9) 153-159

With the rise of the Internet web2.0 and the arrival of the era of big data, NOSQL database has already developed rapidly. The traditional relational database in the treatment of high concurrent Web2.0 pure dynamic website and large-scale data encountered bottleneck. This paper introduces the concept of NOSQL database and data storage model. Taking the HBase database as an example, it describes the system structure and data model of HBase, and it demonstrates data query efficiency of the HBase database which is superior to the relational database. In view of the HBase database storage mechanism, the HBase database and MySQL database for performance, which is proceeding comparison, the result shows that NOSQL database performance in processing large data has obvious advantage over the traditional relational database.

Keywords: data model, NOSQL, HBase, MySQL, performance analysis

Xiaoli Wu, Yating Ying, Feng Zhou Cognitive deviations of information symbols in human-computer interface
Computer Modelling & New Technologies 2014 **18**(9) 160-166

In the design of information symbols in human-computer interface, one meaning usually has several design forms. In order to solve cognitive deviations generated during information exchange, in the paper, we tested the recognition rates of common information symbols. The testing results indicated that users under different cultural backgrounds showed significant differences in information symbol cognition. Users prefer to clear and concise information symbols. Users are inclined to understand the surface meaning of information symbols. Through the study of the recognition rates of typical information symbols, we established the perceptual confusion models of information symbols. Based on different models, designers can improve cognitive deviations of existing symbols and design information symbols, which are consistent with user cognition, for reasonable human-computer interaction.

Keywords: human-computer interface, information symbol, cognitive deviation, perceptual confusion

Yan Zhao, Suyu Dong, Jing Yang Chinese sentiment analysis for commodity price level fluctuation news comments
Computer Modelling & New Technologies 2014 **18**(9) 167-176

With the rapid development of the Internet technology and news media, people pay more attention on news especially about commodity price fluctuation. Hence, more and more Chinese news comments about commodity price fluctuation appear on Internet. These comments contain all kinds of sentiment. Analysing the sentiment of these comments will make government know more about Netizen emotion on this information and enhance efficiency of management, which has important practical significance. In this paper, we adopt three supervised learning methods (naive Bayes, maximum entropy and support vector machines) to automatically classify user comments as two classes (positive and negative). Through a lot of experiments, we found that machine learning techniques perform quite well in the domain of commodity price fluctuation news comments sentiment classification. Meanwhile, the effects of the feature representations and dimensions for the classification of the three machine learning techniques are analysed and discussed in detail. Experimental results show that maximum entropy classifier is best overall. Frequency is a better method of feature representation, which can use fewer features to get better result.

Keywords: sentiment classification, online reviews, supervised machine learning algorithm

Huaping Li Optimal model of highway road based on GPS
Computer Modelling & New Technologies 2014 **18**(9) 177-180

Intelligent transportation system (ITS) has gradually become reality with the increasingly wide application of GPS in

car navigation system. It is wider and wider applied into traffic management to support short-tern traffic prediction, provide the information for public travel rapidly and effectively and realizes the effective supervision on traffic state. It can also relieve the pressure of highway road in our country. This paper focused on the study of floating car data technology based on GPS in ITS and developed relative procedure based on algorithm theory. It designed out a solution suitable for highway real time road condition system based on GPS floating car data technology after the demand analysis, functional decomposition and concept design on highway real time road condition system and best road and shortest time selection on driving by mathematical modelling so that achieve highway road optimization.

Keywords: GPS floating car data, GPS, ITS, mathematical modelling, driving time

Jianping Qiu, Lichao Chen, Guifang Chen Information propagation in social network
*Computer Modelling & New Technologies 2014 **18**(9) 181-185*

Information propagation network analysis provides a new way to investigate online activities. From the perspective of information propagation analysis we can understand it in a constant evolving way, that is, the content of the information is modified by the netizens with a certain probability during the whole propagation process. By analysing the online behaviour of netizens, we constructed an information propagation network on social networks. In this paper, we found that the original information can keep its influence on the netizens only when most of them are forwarders. Meanwhile this paper reveals influence propagation is aggregated, for example, netizens tend to give a low rating after a low rating, as well as a high rating following a high rating. Our findings are helpful in better understanding information propagation.

Keywords: information propagation, social network, scale-free network, power-law distribution, influence propagation

Peng Pan, Xiaojun Cai A complete solution for duplication detection over uncertain data
*Computer Modelling & New Technologies 2014 **18**(9) 186-190*

As the problem of uncertainty for duplication is increasingly prominent with the sharp growth of amount and scale for data sources, we need to pay more attention on it. However, the research on uncertainty about duplicated data is still on its start. In this paper, we propose a complete method for duplication detection with probability, which is efficient and suitable for large-scale dataset. Considering the large-scale background, firstly, we adopt the rapid cluster algorithm based on canopies to get blocks. Secondly, in order to generate the record sets, which represent entity, we provide one fuzzy cluster method over each block by assigning two thresholds. By doing these, we balance the complexity and accuracy. Finally, we assign the probability for each record in one block. The experiments show advantages over other present algorithms for performances.

Keywords: duplication detection, data uncertainty, canopy, data probability

Hsin-Ke Lu, Peng-Chun Lin Towards an organizational learning framework for IS development
*Computer Modelling & New Technologies 2014 **18**(9) 191-195*

Information systems (IS) development has been one of IT managers' concerns for decades. To improve the effectiveness of IS development, many studies argued that development methods need the continuous learning mechanism to better utilize the knowledge along with all development stages. In this study, we discoursed the essential important of organizational learning for development team to deal with the complexities of IS development. The development team failed to learn from the experience in prior IS development projects because of limits of individual learning. Knowledge management concepts are gradually mature and will be fundamental theories to integrate the organizational learning mechanism into IS development. This paper illustrated all related literature and theories and proposed an organizational learning framework for IS development. The framework could aid the development team to recognize the value of organizational learning in the IS development activities and guide for devising the mechanisms for acquiring, maintaining and transferring that knowledge.

Keywords: IS development, organizational learning framework, knowledge management

Wei Li Robust face and facial feature localization using the dual skin model
*Computer Modelling & New Technologies 2014 **18**(9) 196-202*

A fast and adaptive face and facial feature localization algorithm for colour images with sophisticated background is

present. In this algorithm, a self-adaptive pre-processing method was provided to depress the colour bias and the high light. Then the CbgCbr-YIQ dual skin model was proposed to acquire the integrated skin similarity for improving the quality of skin segmentation and extraction. After the morphological post-processing, by using the Adaboost classifier and the information of spatial position, the facial feature positioning was fast realized finally. Experimental results showed the robustness and good performance of the proposed algorithm.

Keywords: pre-processing, dual skin model, facial feature localization, face recognition, Adaboost

Xiaolei Sun, Ning Huang, Jian Zhou, Yue Zhang A novel multi-factor simulation algorithm about tactical network connectivity reliability
*Computer Modelling & New Technologies 2014 **18**(9) 203-206*

Tactical network is a cooperative engagement of a collection of mobile nodes without the requirements of any centralized access points or existing infrastructures. Its connectivity is very crucial. Thus, the calculation of network's connectivity reliability has great significance to further research of tactical network. At present, the study of network is mainly through analytical methods. Those methods build abstract models alone without considering the transmission characteristics and terrain environment of a tactical network. Meanwhile, the existing simulation methods like Monte Carlo, is too abstract and simple. The OPNET simulation can solve the above problems better. In this paper, we analysed the related characteristics of tactical network and designed tactical network connectivity calculating algorithm based on the OPNET simulation. Through the algorithm, we take the transmission and environment features into consideration. As a result, we analysed the effects of wireless transmission characteristics, mobile models and terrain environment on network two-terminal reliability in specific cases.

Keywords: tactical network, OPNET simulation, two-terminal reliability, mobile networking, wireless networks

Jian-Feng Zheng, Ji Zhang Video target tracking with fisher discriminant dictionary learning
*Computer Modelling & New Technologies 2014 **18**(9) 207-213*

As one of the state-of-the-art tracking methods based on sparse coding, l1-tracker finds the target with the minimum reconstruction error from the target template subspace. But the high computational costs restrict its application in practical terms heavily. In this paper, we incorporate the discriminant information into original l1-tracker, and introduce it into the tracking framework, called FD2LT. In our framework, tracking is considered as a problem consisting of object location with dictionary learned in the last frame in generative tracking framework, training samples selection, and dictionary learning with fisher discriminant dictionary learning (FDDL). With our method, the dictionary is much smaller than that in original one, moreover, without loss of tracking performance (and even better in some scenarios). The discriminant power explored from the dictionary is used in generative tracking. Experimental results demonstrate the effectiveness and efficiency of the proposed tracking algorithm.

Keywords: Fisher discrimination dictionary learning, generative and discriminant tracking, sparse coding, video object tracking

Jingfei Cui, Jinlin Wang, Zhen Zhang Implementation of network management software for HINOC
*Computer Modelling & New Technologies 2014 **18**(9) 214-217*

HINOC (High performance Network Over Coax) is Coaxial cable access technology with independent intellectual property rights in China. In the structure of network management software, network topology discovery and maintenance plays an important role in the system. By analysing the traditional algorithms of topology discovery, this paper introduces an algorithm based on SNMP trap and polling. On the basis of test results, the algorithm which has been implemented in HINOC proved to be reliable, efficient and with low-burden.

Keywords: network management, topology management, HINOC, SNMP

Zhiding Chen, Qilun Zou, Qian Liu Evaluating method of DB contracting based cloud model and Gray relational degree
*Computer Modelling & New Technologies 2014 **18**(9) 218-222*

The proposed evaluating method based Cloud Model and Gray Relational Degree aims to solve the fuzziness and randomness problems of evaluation methods and the outcome efficiently. Cloud model is implemented to convert qualitative concept into quantitative value based index system. Gray Relation Degree theory is implemented to access

to evaluation index weights.

Keywords: Cloud Model, Gray relational degree, DB contracting, evaluation of bid, weight

Qiang Fan, Dongjian He, Min Zhang WSN image acquisition method based on interleaving extraction and block compressed sensing
*Computer Modelling & New Technologies 2014 **18**(9) 223-227*

Aiming at disadvantages of current wireless sensor network in the aspect of image acquisition, this thesis proposes a WSN image acquisition method based on Interleaving Extraction and Block Compressed Sensing (IE-BCS). The method uses interleaving extraction to divide an original image into several sub-images at an encoding terminal, then compressive sampling and encoding for each sub-image by means of observation matrix weighted BCS and transmits data to a decoding terminal by their own independent channels. Next, the decoding terminal chooses corresponding decoders according to reception situations and reconstructs the original images by solving sparse optimization problems. Experimental results show that the method can save hardware resources effectively and improve robustness of image transmission.

Keywords: compressed sensing, interleaving extraction, observation matrix, block strategy

Yuansheng Wang, Huarui Wu, Lei Luo, Weihong Ma, Kun Liu Research on cloud storage technologies of typical crop growth environment monitoring data
*Computer Modelling & New Technologies 2014 **18**(9) 228-235*

The typical crop is an important part of regional economy, using of information technology tools to enhance the fine management of typical crop, improve the efficiency and level of agricultural production will play an important role to promote regional economic development. The Internet of things technologies can provide conditions for precise management of crop, but it will also lead to new technical application difficult problems, focusing on the storage and processing problem of the "real time, high frequency, mass, rapid growth" monitoring data of the crop growth environment, this paper put forward a network management and scheduling method for massive data based cloud storage technologies to solve the large data storage and concurrently access bottlenecks of the traditional relational storage, and provided good technical conditions for the typical crop production management.

Keywords: typical crop, growth environment monitoring, cloud storage

Wangcheng Cao, Lin Zheng, Kuiling Dong Image adaptive filtering based on the improved Alpha-trimmed mean algorithm
*Computer Modelling & New Technologies 2014 **18**(9) 236-239*

Filtering is an important research direction of image processing. In view of the characteristics of image noise, an adaptive image filtering algorithm is proposed based on the improved Alpha-trimmed mean algorithm. The algorithm dynamically selects parameter d of Alpha-trimmed mean algorithm through the calculation of the pixel correlation within the filter window, so that the algorithm can filter accordingly with the degree of noise. The experiment results show that the filtering effect of the proposed algorithm is excellent for the images corrupted by Gaussian noise, salt and pepper noise or mixed noise, and this algorithm is also capable of maintaining the detailed information of the original image.

Keywords: image filtering, alpha-trimmed mean filter, neighbourhood correlation

Operation research and decision making

Lei Ren, Stephen Nash, Michael Hartnett Optimal interpolation data assimilation of surface currents by utilizing pseudo measurement with Monte Carlo simulation
*Computer Modelling & New Technologies 2014 **18**(9) 240-248*

Optimal Interpolation (OI) data assimilation is a technique to combine available observations with background states to improve prediction states. In this research, pseudo measurement of surface currents generated by adding noise with Monte Carlo simulation is used to update the background states with optimal interpolation. The core of Optimal Interpolation data assimilation is the definition of background error covariance, which determines to what extent the model background states will be corrected to match the observations. The background error covariance is computed

before the data assimilation process. The model background errors are calculated from the mean over a short time interval ten minutes. A series of sensitivity tests with Optimal Interpolation are done by calculating Root Mean Square Error (RMSE) to decide the appropriate parameters. The improvement of Optimal Interpolation at reference points is measured in Taylor diagrams, and the surface current maps of test domain show the effectiveness of Optimal Interpolation.

Keywords: Optimal Interpolation, data assimilation, Monte Carlo, pseudo measurement, background error covariance

Yanwen Wang, Xiuju Gao, Liming Yang Logistics outsourcing and selecting of logistics service provider of the e-commerce companies: a fuzzy TOPSIS approach
*Computer Modelling & New Technologies 2014 **18**(9) 249-255*

Logistics outsourcing has become the development trend of enterprise logistics operations. A good logistics service provider can improve the customer satisfaction, as also as reducing the cost of the whole supply chain, so it is very important to evaluate the logistics outsourcing service for the corn companies in the supply chain. This paper is an attempt to identify the main factors of selecting satisfactory logistics service provider of the electronic commerce (e-commerce) companies in China. GRA and Fuzzy TOPSIS are employed to evaluate the service of the logistics companies. The managerial implications are also discussed in the last section.

Keywords: logistics outsourcing, e-commerce, Fuzzy TOPSIS, evaluation

Dong-min Xu, You-fang Huang, Bin Yang Model study on stochastic flow logistics network considering neighbourhood information of nodes
*Computer Modelling & New Technologies 2014 **18**(9) 256-264*

The significance and stochasticity of node have an effect on logistics network activity. This paper describes the importance of node by integrating the node neighbourhood information, which relates to the neighbourhood arc energy consumption and node degree which shows the node connectivity in logistics network, and during which this paper introduces the parameter α to represent the preference degree of energy consumption. This paper develops a multi-objective model seeking to energy consumption minimization and reliability maximization in the light of node stochasticity and the influence of node importance. The MPs (Minimal Path Set) is employed during the process of modelling. Genetic algorithm is applied to solve this model and a numerical experiment is presented to demonstrate the effect of significance and stochasticity of node to logistics path choice and flow assignment.

Keywords: stochastic flow network, neighbourhood information, energy consumption constraint, multi-objective model, logistics network

Yang Zhao, Jie Tan Research on the relationship between knowledge transfer models to various strategic alliances
*Computer Modelling & New Technologies 2014 **18**(9) 265-271*

As a part of knowledge management research, this paper focuses on knowledge transfer within strategic alliances and attempts to classify the basic models of knowledge transfer that actually take place within SA so that get to know on how partners chose the most suitable knowledge transfer model in SA. This paper presents four kinds of knowledge transfer model in strategic alliances by summing up the outcome of research on this issue. Based on the conclusion, this paper then discusses the correspondence between the types of the strategic alliances and choosing these knowledge transfer models and five corresponding propositions about the issue is given at last.

Keywords: knowledge management, strategic alliances, knowledge transfer, tacit knowledge, explicit knowledge

Qiang Song, Ai-min Wang Power transformer diagnostic prediction research based on quantum neural networks and evidence theory
*Computer Modelling & New Technologies **18**(9) 272-276*

Aiming at the fault of power transformer fault information diversity and uncertainty, a large amount of data and no regularity characteristics, a new fault diagnosis method of quantum neural network based on information fusion. In order to accurately and effectively identify transformer fault model, combining the quantum neural network and evidence theory combination of transformer fault diagnosis. A quantum neural networks to collect data on the

macroscopic, microscopic quantum corrections in the interval of fuzzy intersection data according to a certain proportion of the rational allocation of the associated mode, so as to improve the accuracy of pattern recognition; use of the evidence theory can improve the convergence speed of quantum neural networks. The results were compared with the diagnosis and BP neural network input, that this method has a higher accuracy in transformer fault pattern recognition.

Keywords: transformer diagnostic prediction, combination forecasting, quantum neural networks, evidence theory

Nan Ma, Yun Zhai, Bingru Yang Survey of research directions in fuzzy cognitive map
*Computer Modelling & New Technologies 2014 **18**(9) 277-283*

Fuzzy Cognitive Map as a generally recognized intelligent tool has been widely used in the data mining and the machine leaning fields. This paper reviewed the present research situation of Fuzzy Cognitive Map at home and abroad in recent years. Firstly, it reviewed the development progress of Fuzzy Cognitive Map with its advantage over other fuzzy learning methods, followed by the research papers indexed by EI database, ACM database and the SCI database. Then it summarized the classification and the learning methods of Fuzzy Cognitive Map. Furthermore, the applications of Fuzzy Cognitive Map are analysed finally. We hope our work help the people to have a general understanding and push the Fuzzy Cognitive Map forward.

Keywords: fuzzy cognitive map, development, classification, learning methods

Lin Li Exploring the relationship between inventory level and bullwhip effect in the supply chain
*Computer Modelling & New Technologies 2014 **18**(9) 284-289*

This paper intends to work out an expected inventory level formula for the retailer in the two-stage supply chain. It aims at disclosing the quantitative relationship between bullwhip effect and expected inventory level and does analysis of simulation experiment. The model supposes the market demand faced by the retailer follows the autoregressive process AR(1) and that the retailer makes anticipation of the market demand during the replenishment lead time by mean square error method. Moreover, if the interference factor follows normal distribution with mean 0 and variance σ^2, market demand in every period, demand estimation during the replenishment lead time and the order quantity made by the retailer are all proved to follow normal distribution.

Keywords: bullwhip effect, first-order autoregressive model, supply chain management, inventory level, demand forecasting

Chengduan Wang A research into location routing problem based on hybrid genetic simulated annealing algorithm
*Computer Modelling & New Technologies 2014 **18**(9) 290-297*

With the diversified and personalized commodity requirements as well as small batch dispatch and frequent dispatch features under the circumstance of E-commerce, the environment for logistics dispatching system becomes increasingly complicated and the inter-influence between different subsystems in logistics system optimization becomes increasingly significant. As a result, judged from the aspect of logistics system integration, after taking customer's personalized dispatching time into consideration, it's necessary to establish location routing problem with changeable softtime windows model. Based on the feature of the model, this paper adopts hybrid genetic simulated annealing algorithm to gain solution. The experimental result shows that this algorithm is much better than gaining solution solely by adopting hybrid genetic algorithm or simulated annealing algorithm in the aspect of optimal solution, solution quality, calculating efficiency and algorithm stability.

Keywords: hybrid genetic simulated annealing algorithm, logistics system, location routing problem

Hongxia Wei Research on computer network English language learning
*Computer Modelling & New Technologies 2014 **18**(9) 298-302*

Network information technology affects English language learning mode all the time with the rapid development of information technology. University English course teaching also turns from traditional classroom to modern network teaching. As a bran-new teaching form, university English teaching brings not only advantage for classroom teaching but also impact on teaching idea, teaching means, teaching pattern, etc. It breaks through the ecological balance of traditional university English course and affects the successful implementation of teaching reform mode. Aiming at improving university English course teaching quality and perfect teaching reform mode, this paper studied new English

teaching means based on computer network technology, analysed imbalance of university course teaching in teaching reform mode and then strove to construct a frame of ecologicalization university English course teaching.

Keywords: network environment, university English teaching, voice recognition confidence level, digital game teaching

Lizheng Liu, Fangai Liu, Feng Yang A research about the predictive control of dynamic feedforward neural network based on particle swarm optimization
*Computer Modelling & New Technologies 2014 **18**(9) 303-311*

The paper proposes the Dynamic Feedforward Neural Network based on Hidden Particle Swarm Optimization (HPSO-DFNN) to deal with the model predicative control problem of unknown nonlinear delay systems. It realizes quick, precise system modelling for controlled objects. Besides, the Smith predictive double controllers are designed to separate fixed set point control from external disturbance. The DFNN based on large-scale PSO is treated as an identifier and a predictor for the complex controlled objects with the purpose of increasing the robustness of the control system. Furthermore, aiming at the problem of constrained multi-input-multi-output (MIMO) model predictive control, rolling optimization is conducted to obtain controlled quantity through the PSO algorithm. After that, a combined neural network structure is put forward and applied to system modelling. Finally, the paper uses the typical nonlinear model to verify its effectiveness.

Keywords: nonlinear delay systems, dynamic feedforward neural network, particle swarm optimization

Min Liu, Jing Cao, Yanru Xue, Yinghua Yao, Xuezuo Zhao Development and research on remote online education information system based on Web
*Computer Modelling & New Technologies 2014 **18**(9) 312-318*

Remote education is a way of developing education-teaching activities, which transmits outstanding teaching resources to massive students scattered in different time and space by the Internet, satellite and other methods of communication technologies. The Internet is the most convenient way to participate in distance learning. This paper applies B/S three-layer architecture and ADO data access technology to complete the design of the remote online education information system based on Web. The system according to the actual business needs of remote education, carry on the design based on the role of application and division of the task, and it composed of modules of courseware on demand, online management, online examination, online exercise, teacher management, student management, etc.

Keywords: remote online education, information system, Internet

Shunye Wang A research on the intelligent multi-objective optimization problem based on wavelet theory and neural networks
*Computer Modelling & New Technologies 2014 **18**(9) 319-328*

Aiming to solve the multi-objective optimization problem caused by wavelet multiresolution a1nalysis (MRA), the thesis improves the original multi-objective non-dominated genetic algorithm. After fast non-dominated sorting, the evolution of population is achieved through particle swarm optimization (PSO). In this way, the thesis realizes a more effective, organic combination of the multi-objective optimization problem and neural networks. MRA is a natural fit for the multi-objective optimization problem. The ability of neural networks to deal with complex errors is improved through error decomposition based on different wavelet decomposition scales.

Keywords: multi-objective optimization problem, wavelet theory, PSO, neural networks

Yulan Zheng Design and algorithm of supply chain network model based on uncertain environment
*Computer Modelling & New Technologies 2014 **18**(9) 329-333*

Under the conditions of modern market economy and the increasing pressure of competition, a growing number of product life cycles are getting shorter, customer demand is constantly changing, this force the companies began to pay attention to their own internal management and supply chain management. Supply chain refers to the suppliers, manufacturers, distributors and end-users, and the network, which design both directions of material flow, information flow and capital flow [1]. In order to manage and use it better, People put network information technology in supply chain management, developed or purchased ERP implementation systems software, thus can share information between departments and enterprises. But problems have gradually exposed, faster access and processing does not produce a

good decision automatically. So this paper under the premise of uncertainty, started studying supply chain's network model, establishing optimization model, design algorithms, therefore reducing the total costs, to avoid risks.

Keywords: uncertain environment, supply chain, model, design

Zhiming Wang, Shuzhen Zhu An empirical study on China listing corporation industrial-financial combination based on the regression method
*Computer Modelling & New Technologies 2014 **18**(9) 334-338*

This paper takes Chinese listing corporation as research samples to find the best combination area or point about the industry-finance combination on enterprise operating performance. Research on the combination of behaviour impact is carried out through EPS, RPE, ROE, Tobin Q four indicators. The paper is mainly using the nonlinear regression method based on steel listing corporation annual data from 2005 to 2008. Experiments show that the relation by EPS, RPE and Tobin Q meeting the cubic function curve is significant; the relation by ROE meeting the curve function is not significant. In a certain stage of development of the enterprise the appropriate industry-finance combination area or ratio exists, but the ratio of each index is different.

Keywords: industrial-financial combination, the nonlinear regression, the cubic curve, China steel listing corporation, the best combination point

Xiuli Li, Yuhong Zhang, Sujuan Zhao Research on grey correlation analysis model of enterprise human resources competitiveness
*Computer Modelling & New Technologies 2014 **18**(9) 339-344*

Human resources is one of the key indexes in measuring the competitiveness of en enterprise. However, evaluation of enterprise human resource competitiveness, evaluation index system and evaluation model remain a problem in academic and practice. Nowadays, key factors in evaluation of enterprise human resources competitiveness and evaluation index are incomplete and imperfect, and processing methods of corresponding evaluation index is not scientific. Thus, studies on analysis model of enterprise human resource are of great importance. In this paper, an improved grey correlation analysis model of human resource competitiveness was put forward. This model provides relative restrictive factors in analysis of enterprise resource competitiveness, and analysed enterprise competitiveness from the aspects of human resource quality competiveness, market development competiveness, management quality competiveness, etc. and evaluation analysis based on improved grey correlation analysis method was conducted. Its evaluation result can be a basis of selection of human resource developing strategies for direction and of frame of decision-making. Finally, the model and algorithm was proved feasible by implementation of actual case.

Keywords: human resources, competitiveness, grey correlation analysis, evaluating indicator, model

Qiaojuan Feng, Yongjiao Wang A membership degree algorithm of collaborative design roles in distributed design transaction
*Computer Modelling & New Technologies 2014 **18**(9) 345-349*

As there is a priority in multiple design transaction and multiple design roles in the distributed design traction, this paper studies the distribution hierarchy of collaborative design roles and proposes a membership degree algorithm based on grey related analysis. In this algorithm, design constraint degree about the design roles is acquired through grey related analysis. Under different design constraint degree, design roles and incidence coefficient are acquired together with corresponding matrix. Then the membership degree can be available by the constraint degree and the matrix. Empirical test proves the efficacy and practicability of this algorithm.

Keywords: distributed design, collaborative design, membership degree, algorithm, artificial intelligence

Zhenyu Yang Actor-Critic reinforcement learning based on prior knowledge
*Computer Modelling & New Technologies 2014 **18**(9) 350-357*

In order to improve the incremental learning algorithm Actor-Critic learning efficiency, from a policy learning, introduce experience sample data into incremental Actor-Critic algorithm, make effective use of the useful information contained in the sample data of experience in the learning process. Given the recursive least-squares temporal difference, RLSTD (λ) algorithm and incremental least-squares temporal difference, iLSTD (λ) algorithms are able to

make good use sample data collected in the past, respectively RLSTD and iLSTD algorithm is applied to policy evaluation Critic's. Then, Critic learned value function based on RLSTD or iLSTD algorithm, Actor gradient update strategy based on conventional parameters, so the improvement of Critic effectiveness assessment will help Actor to improve strategy-learning performance. Finally, simulation studies on two control problems with continuous state space, analyse the impact of different parameters on the performance of the learning algorithm and verify its effectiveness.

Keywords: actor-critic, RLSTD, on-policy, ILSTD

Kehe Wu, Yi Li, Long Chen, Fei Chen Research of deep packet inspection technology in industry control systems based on d-Left counting bloom filter
*Computer Modelling & New Technologies 2014 **18**(9) 358-364*

Aims at the problem of industry control system logic being tampered by malicious programs, this paper proposed a deep packet inspection model for industry control systems. It adopts d-Left Counting Bloom Filter to implement string matching of characteristic database, which has a higher matching efficiency and lower misjudgement rate compared to other algorithms. This model can monitor malicious behaviours in industry control systems effectively and can basically meet the demand of security and reliability in current industry control systems.

Keywords: industry control system, deep packet inspection, d-left counting bloom filter, information security

Tianyan Wu, Jianjun Zhan, Wei Yan Study on configuration sequence of indemnificatory community public service facility based on MIV-BP Neural Network
*Computer Modelling & New Technologies 2014 **18**(9) 365-370*

China is seeing large-scale construction of indemnificatory community being. Yet due to lack of dynamic planning and arrangement in advance and little consideration of public service facility configuration sequence, the configuration of public service facility in indemnificatory community is lagging behind and inefficient, failing to attract the residents to move in the community. This paper structures the MIV-BP Neural Network Model, and gives an empirical analysis on the influence sequence of the indemnificatory community public service facilities to the population occupancy rate. The results suggest that the configuration of public service facility in indemnificatory community should be sequentially configured in period and in grade according to the community's specific present situation and developmental conditions as well as the continuous increase of population occupancy rate.

Keywords: indemnificatory community, public service facility, configuration sequence, MIV-BP neural network

Caixian Chen, Huijian Han, Zheng Liu KNN question classification method based on Apriori algorithm
*Computer Modelling & New Technologies 2014 **18**(9) 371-379*

KNN (K-Nearest Neighbours) algorithm is a classification algorithm that can apply to question classification. However, its time complexity will increase linearly with the increase of training set size, which constrains the actual application effects of this algorithm. In this paper, based on a discussion of disadvantages of traditional KNN methods, an improved KNN algorithm based on Apriori algorithm was proposed. This method extracts the frequent feature set of training samples of different categories and the associated samples. Next, on the basis of correlation analysis of each category of samples, a proper nearest neighbour number k was determined for an unknown category of questions. In the training samples of known categories, k nearest neighbours were selected. And then, according to the category of nearest neighbours, the category of unknown question was identified. Compared with the question classification method of traditional KNN, the improved method could efficiently determine the value of k and decrease time complexity. Our experimental results demonstrated that the improved KNN question classification method improved the efficiency and accuracy of question classification.

Keywords: question classification, KNN, correlation analysis

Qin Yang, Changyao Zhou Cost-Sensitive learning on classification
*Computer Modelling & New Technologies 2014 **18**(9) 380-386*

Real-world predictive data mining (classification or regression) problems are often cost sensitive, meaning that different types of prediction errors are not equally costly. In this paper we propose a new algorithm for cost-sensitive classification in a multiple time series prediction problems. The fitness function of the genetic algorithm is the average

cost of classification when using the decision tree, including both the costs of tests (features, measurements) and the costs of classification errors. The proposed model is evaluated in a real world application based on a network of satellite network map distributed in land spatial pattern evolution in Chengdu Plain. These satellite networks generate multiple time series data representing land spatial pattern. This study presents a new algorithm for cost-sensitive classification that deal with class imbalance using both recompiling and CSL. The method combines and compares several sampling methods with CSL using support vector machines (SVM). We build our cost-benefit model for the prediction process as a function of satellite network in a distributed land spatial and measured the optimum number of satellite network that will balance the expenses of the system with the prediction accuracy.

Keywords: cost sensitive learning, prediction, distributed satellite network, cost benefit analysis

Shenyi Qian, Yanling Zhu, Shen Li A medical quality evaluation method based on combined weight
*Computer Modelling & New Technologies 2014 **18**(9) 387-391*

Medical quality evaluation is the key and important link of the current medical institutions improve the core competitiveness, considering the characteristics of the medical industry, the paper constructs the surgical and non-surgical medical quality evaluation index system. In addition, the traditional medical quality evaluation in determining the index weight coefficients are too single, it is easy to cause the subjective assessment results too much or the accuracy is not high. Therefore, the paper using the Analytic Hierarchy Process (AHP) for subjective weight and using the Rough Set (RS), then get them together and put forward an approach of medical quality evaluation method based on combined weight, this method absorbs the advantage of them and overcomes the disadvantages of them and achieve the complementary advantages. Finally through the case analysis, verifying the feasibility and effectiveness of the method.

Keywords: medical quality evaluation, indicator system, combined weight, AHP, RS

Xuejing Du, Huanhuan Guo, Zhanyu Wang Research of body dimension of Chinese adult male and application
*Computer Modelling & New Technologies 2014 **18**(9) 392-397*

The goal of this study is to obtain accurate human dimension of current Chinese shape. Anthropometric data of adult male in GB10000-88 is adopted as sample observation data in this paper. Based on similarity theory of human body and the correlation of human body static dimensions, by statistical analysis of the sample date, the regression equation is established which determines the relationship between measurement of each part of the male body and height, weight. The maximum measurement error is within 5% by comparing the measurement data with calculating dimension data obtained by regression equations established in the paper, which verifies the rationality and accuracy of the regression equations. Once parameters of height and weight are given, the regression equations can quickly and easily provide effective body dimensions data used to build crash simulation model specifically for Chinese and man-machine product.

Keywords: automobile crash, dummy model, human dimensions, occupant restraint, regression equation

Xuanxuan Xu Decision to rescue failed product innovation projects based on the leading innovation strategy
Computer Modelling & New Technologies 2014 18(9) 398-403

The leading innovation strategy is one of the important methods used to rescue failed product innovation projects. Rescuing failed projects based on the leading innovation strategy, rescuing time and points of competitors participating in the market can affect the probability of success. The established mathematical economic model can be used to analyse the cost of rescuing failed product innovation projects and income when competitors participate in the market in the introduction period, growth period, or mature period. The model can also be used to determine the feasibility of the leading innovation strategy and measures that enterprises should take to obtain greater profits under different competitive environments.

Keywords: leading innovation, product innovation, failed projects, rescue

Jian-gang Shi, Jin-can Liu Mechanism of formation in the selling price of urban industrial land in China based on the multi-attribute auction theory
*Computer Modelling & New Technologies 2014 **18**(9) 404-409*

There is the fierce competition for investment introduction among local governments in China. Thus, the auction relationship about urban industrial land belongs to a kind of reverse auction between local government and investment enterprise. To explore the mechanism of formation in the selling price of urban industrial land, we build a multi-attribute first-score sealed-bid auction model. Our studies reveal that the selling price of urban industrial land is the decreasing function of local government's expected comprehensive earnings from the investment introduction, the intensity of competition for investment introduction among local governments or its cost factors of providing some attributes. However, it is the increasing function of enterprise's weights for different attributes. They appropriately explain why the selling price of urban industrial land has been at the low level for a long time in China. Yet, the comprehensive earnings and the realization of land asset value are not a pair of irreconcilable conflicts. Local governments should prioritize the attributes, which the investment enterprises care about, then reduce the cost of providing these attributes. In the end, it will resolve the interest conflicts in the process of industrial land operation.

Keywords: industrial park, industrial land, selling price, multi-attribute auction

Dai Linlin Statistical analysis and prediction of Qingdao urban consumer price index
Computer Modelli Ng & New Technologies 2014 18(9) 410-414

In recent years, Qingdao faces inflationary pressure along with the fast development of economy. The urban consumer price index (CPI) is an important measure of regional inflation level. Thus, analysing the influencing factors of the CIP and determining the key factors is necessary to supply scientific evidence for the economic management departments to adjust the industrial structure. This paper filtrates and eliminates six indexes likely to affect urban CPI of Qingdao based on Studentized residual, correlation analysis, and multiple collinearity diagnosis. This paper also performs a statistical analysis of the urban CPI of Qingdao from 2005 to 2012 and draws the conclusion that food price is the largest index that affects CPI. Basing on this conclusion, corresponding prediction models are established to supply statistics for departments that adjust the economy for the future.

Keywords: Studentized residual, correlation analysis, collinearity diagnosis, multiple linear regression model

Junting Lin, Xiaoming Wang, Jianwu Dang Reliability and safety verification of the new collision avoidance strategy for Chinese train control system
Computer Modelling & New Technologies 18(9) 415-422

Although equipped with comprehensive and complex technologies on the railway to avoid collisions between trains, such as train control system extensively deployed in the infrastructure, there are still enormous amount of collision between trains. The safety of railway operation mainly depends on the correct operation of the interlocking and train control system, because only the control centre has an overall overview of the traffic situation, and a train driver could only be warned of a possible collision if the operation centre decides it. Experiences from the aviation, the maritime navigation and road transportation have shown that the probability of collisions could be significantly reduced with infrastructure-less collision avoidance system. In this paper, an enhanced safety strategy, namely, Collision Avoidance System overlay Train Control System (CASOTCS) for Chinese railway was provided, which was based on direct vehicle-to-vehicle communication and independent of the regular control mechanism. CASOTCS can receive and evaluate the information broadcasted by other neighbouring trains all the time, which will trigger collision alert and provide a solution for the collision when a potential collision is detected. The unit architecture and its key issues of CASOTCS were also discussed. After the analysis of reliability and safety of both the traditional train control system and the enhanced CASOTCS, the reliability block diagram and isomorphic Markov model were established in the paper. Through the calculation of the indexes of the reliability and safety about the two kinds of control systems, a significant conclusion has been made, that is, the reliability and safety of the train control system plus collision avoidance system are much higher than that of the traditional one and CASOTCS has the ability to increase the reliability, safety and efficiency in the future.

Keywords: Collision Avoidance, Chinese Train Control System, Safety Verification, Reliability, Markov

Yun-jun Yu, Yun-tao Xue, Sui Peng, Chao Tong, Zi-heng Xu Short-Term photovoltaic system power forecasting based on ECSVM optimized by GA
Computer Modelling & New Technologies 2014 18(9) 423-429

It is of great significance to research PV forecasting techniques for mitigating the effects of the randomness of the Photovoltaic output. This paper analyses many factors from PV which impact photovoltaic output and extracts the main factors, forming sample data combined with the historical database generation data from PV monitoring system. And an error correction SVM method (ECSVM) is used to calculate the open integration of photovoltaic power storage system in advance or after the time in order to try to eliminate the system error between the predicted and actual values. At the same time, using genetic algorithm to optimize kernel function parameter and the error penalty factor and other parameters in this model, the establishment of the GA-ECSVM model improves portfolio optimization model parameter prediction accuracy and efficiency of the selected type. Finally examples verified and compared with standard SVM methods and ECSVM method, predicting effects show that: The GA-ECSVM optimization model presented in this paper has better learning ability and generalization ability in the short term prediction of photovoltaic power generation, with the prediction accuracy of 95.2016%.

Keywords: PV, SVM, error correction, GA, forecast

Wensu Xu, Leilei Wang Research on the application value of computer technology in hotel management
Computer Modelling & New Technologies 2014 18(9) 430-434

Hotel management become much important with the increase of services of hotel. And the traditional hotel management pattern have not adapted to the modern hotel management. Therefore, management with high-tech technology is needed in modern hotel. Computer application can well solve management mode of modern hotel and improve the service level of hotel. With the development of IT technology and hotel industry, current hotel management have developed into a new age and new systematic development direction emerges constantly. This paper mainly discusses the application value of computer technology in hotel management. And it illustrates the importance of computer technology in hotel management by explaining the systematic design and implement of hotel management.

Keywords: hotel management, computer, information system

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

N P Pravednaya, S F Baranova Fractal model of electric conductivity of aging Al alloys
Computer Modelling & New Technologies 2014 18(9) 435-437

Theoretical study of anisotropy changes of crystal structure in Al alloys as a result of senescence is performed. The appropriate model is developed that allowed to explain experimental results concerning the influence of tempering and cold rolling on the anisotropy of electrical conductivity of Al alloy D16.

Keywords: crystal anisotropy, plastic deformation, diffusion, dislocations, electrical conductivity list

Wei Li Gu, Han Qing Wang, Guang Xiao Kou, Qiao Ying Cao The finite time thermodynamics analysis and the energy-saving optimization of the coil organic heat transfer material heater
Computer Modelling & New Technologies 2014 18(9) 438-444

According to the problems as low efficiency, aging of organic heat transfer material and frequent accidents in the operation of the coil organic heat transfer material heater, with the finite time thermodynamics, this paper studies the actual processes including combustion process, flow and heat transfer process, considers the irreversible factors as combustion condition, the fuel characteristics, flow resistance and heat transfer temperature difference, derives the calculation formula of energy loss for three processes and to determine the minimum energy loss. Based on this, this paper proposes energy-saving optimization strategy and gives examples, the results accord with the analysis. In design and operation, the outlet flue gas temperature of furnace, flue gas temperature and flow velocity of hot oil are important.

Keywords: the coil organic heat transfer material heater, finite time thermodynamics, energy loss, energy-saving optimization

Chunyu He, Zhijie Jiao, Di Wu Plastic coefficient on-line calculation method for hot rolling
Computer Modelling & New Technologies 2014 18(9) 445-449

Aiming at the requirement of the high precision rolled piece plastic coefficient in the hot rolling process, this paper puts forward an on-line calculation method for the plastic coefficient. Based on the rolling mechanism model and fitting the plastic curve by the quadratic curve, the paper is to solve the tangency at the rolling pressure point, which is the plastic coefficient. The plastic coefficient calculation method exploited in the paper could be embedded in the basic automation directly before the calculation of the automatic gauge control (AGC) regulation quantity. According to the changes of the rolling force and the roll gap, the method can make the correction computation on the plastic coefficient, so as to

improve the thickness compensation accuracy of the AGC system.

Keywords: hot rolling, plastic coefficient, quadratic curve, AGC

Bo Jin, Lijun Zhao, Shiqiang Zhu Error modelling of depth estimation based on simplified stereo vision for mobile robots
*Computer Modelling & New Technologies 2014 **18**(9) 450-454*

Depth estimation is the precondition in obstacle avoidance for mobile robots. To improve the obstacle detecting effectiveness and quickness in poor-textured backgrounds, we used the centroid abscissa difference of corresponding obstacle region in image pairs as parallax to estimate obstacle depth. The error of parallax and depth were studied analytically and numerically. Wood blocks of different shapes and sizes were used for demonstrating the relationship between estimated depth and actual depth. A quadratic function model was obtained after experiments. Although the depth estimation error was relatively higher compared to conventional grayscale correlation-based method, the proposed method was expected to satisfy the accuracy requirement of depth estimation for common mobile robots.

Keywords: error model, depth estimation, stereo vision, radial distortion

Wei Liu, Hongzhao Liu A study of characteristics extraction of dynamic pressure signals in pipeline based on EMD
*Computer Modelling & New Technologies 2014 **18**(9) 455-461*

Characteristic selection is a key to accurate signal recognition. Specific to the signal recognition of dynamic pressure waves in pipelines, this article proposes to use EMD method to decompose dynamic pressure signals into a series of IMFs, adopt the correlation theories of signals to determine and eliminate noise, retain valid IMFs from 4th to 8th level and extract kurtosis, energy and statistical characteristics. The experiment analyses and verifies that the energy or statistical characteristics (mean variance, mean standard deviation, mean range and mean quartile range) can effectively represent signals.

Keywords: EMD, kurtosis, energy, statistics, characteristic

Kaijun Wu, Tianqi Luo, Huaiwei Lu, Yazhou Shan Dynamical behaviour and coupling synchronization of oscillatory activities in a cell system of neural network
*Computer Modelling & New Technologies 2014 **18**(9) 462-467*

The paper mainly described a mathematical model of calcium ion oscillation of non-excitable cells. Based on the model, complex oscillations caused by variations of bifurcation parameters were analysed in detail and effects of variation of parameters to synchronization were discussed when the coupling intensity of two gaps-junction- coupled calcium ion cells was certain. Through numerical simulation, under less stiffness of coupling and among a certain scope of parameters, the phenomenon that the larger the parameters were, the easier the system occurred approximate synchronization were further illustrated.

Keywords: cluster oscillation, bifurcation, fast-slow dynamics, synchronization

Chunhun Liang, Yuanchun Li Finite time control for probe descent near small bodies
*Computer Modelling & New Technologies 2014 **18**(9) 468-476*

For the purpose of probe soft landing on small bodies safely, this paper focuses on the orbital dynamics and describes a new control algorithm for the probe descent near a rotating small body. The general formulation of the probe equations of motion in the Body-Fixed Coordinate system is obtained through Newton's second law firstly. Then the nominal polynomial trajectory of vertical direction is planned in the condition of fuel consumption suboptimum. Considering uncertainties and perturbations, the control laws based on Adaptive Terminal Sliding Mode with compensation term are developed to track the desired trajectory finally. Suppose the initial conditions presented in this paper, position and velocity errors tend to zero in the finite time in the phase of sliding mode motion so it will make the task succeed during descent phase near small bodies. Finally, the effectiveness of the guidance and control algorithm is verified by MATLAB simulations. The proposed algorithm can fast and accurately track the planned trajectory in the finite time and is robust to parameter uncertainty, feedback state error and external disturbances. The validity is conformed by

computer simulation.

Keywords: adaptive terminal sliding mode control, soft landing, descent near small body, guidance and control

Hongyang Zhang Research on modelling of intake tower in three-dimension CAD software and simulation analysis in FE software
*Computer Modelling & New Technologies 2014 **18**(9) 477-483*

ANSYS is one of the most influential finite element analysis software in the world because of its very powerful calculation and analysis ability, but its pre-process function is weak relatively. SolidWorks is the three-dimension parametric feature modelling software of 100% feature modelling and 100% parameterization, which provides product-level automated design tools. In this paper, combining with the intake tower, it discusses the method of modelling in three-dimension CAD software SolidWorks and the interface processing between SolidWorks and the ANSYS code, which decreases the difficulty in modelling complicated models in ANSYS. In view of the function of the birth-death element and secondary development with APDL (ANSYS parametric design language), simulation analyses of thermal field and stress during the construction and impounding periods were conveniently conducted.

Keywords: computer modelling, SolidWorks, interface processing, birth-death element, APDL, simulation analysis, ANSYS

Shuang Du, Chuncheng Zuo Application of support vector machine in driving ranges prediction of pure electric vehicle with dual – energy storage system based on particle swarm algorithm
*Computer Modelling & New Technologies 2014 **18**(9) 484-490*

Driving ranges was a key factor that may affect the popularization and development of pure electric vehicle (PEV) with dual-energy storage system (DESS). It relied on neural network for its prediction. However, the prediction effect was not satisfactory due to local minimization, slow convergence rate, overfitting phenomenon and so on. In order to be more accurate in prediction, this paper introduced the Support Vector Regress (SVR) to the vehicle with parameters optimized by particle swarm optimization (PSO). Compare to BP neural network algorithm, PSO-SVR algorithm is more accurate and practical.

Keywords: pure electric vehicle with dual-energy storage system, particle swarm optimization algorithm, support vector machine, driving ranges

Xiaoheng Yan, Hua Fu, Weihua Chen Prediction of coal mine gas emission based on Markov chain improving IGA-BP model
*Computer Modelling & New Technologies 2014 **18**(9) 491-496*

There are a lot of factors that affect the gas emission, and among those there is a complicated and nonlinear relationship, so a BP neural network model based on immune genetic algorithm (IGA) was constructed to solve the problem of the traditional BP neural network such as, slow training speed, easy to be trapped into local optimums, and the premature convergence. In order to further improve accuracy of the prediction, the Markov chain was used to modify the residual series for the sample of bigger error. The correction result is more close to the measured value. The results showed that both the prediction accuracy and convergence speed of the IGA-BP model are better than the BP neural network model. The prediction after modified by Markov chain was further improved, the absolute average relative error of the prediction of the IGA-BP model is 2.40%.

Keywords: gas emission, immunity genetic algorithm, BP neural network, Markov chain, prediction

Zhongshuang Wang, Yangyang Tao A computer aided kineto-static analysis method for spatial robot mechanism based on vector bond graph
*Computer Modelling & New Technologies 2014 **18**(9) 497-502*

In order to increase the reliability and efficiency of the kineto-static analysis of complex robot systems, the corresponding vector bond graph procedure is proposed. From the algebraic relations of input and output vectors in the basic fields, junction structure and Euler-junction structure of system vector bond graph model, the unified formulae of driving moment (or force) and constraint forces at joints are derived, which are easily derived on a computer in a complete form. For the algebraic difficulties brought by differential causality and nonlinear junction structure in system automatic modelling and kineto-static analysis, the effective bond graph augment method is proposed. Based on the

kinematic constraint relations, the vector bond graph model of the spatial robot mechanism with five degrees of freedom can be made. As a result, the automatic modelling and kineto-static analysis of complex robot system on a computer is realized, its validity is illustrated.

Keywords: robot mechanism, kineto-static analysis, vector bond graph, causality, joint constrain

Zhongshuang Wang, Zouqi Shu The dynamic modelling and simulation for vehicle suspension systems based on vector bond graph
*Computer Modelling & New Technologies 2014 **18**(9) 503-509*

A dynamic modelling and simulation procedure for vehicle suspension system based on vector bond graph is described. According to kinematic constraint relations of components, the vector bond graph model for vehicle suspension system is built. In consideration of the coupling of energy variables and coenergy variables in independent energy storage field and dependent energy storage field. The unified formulae of system state space equations which are easily generated on a computer is derived. As a result, the automatic modelling and simulation of dynamics for vehicle suspension system can be realized based on MATLAB.

Keywords: vector bond graph, vehicle suspension system, modelling and simulation, mixed causality

lei Wang Optimal design of machine tool bed based on Ansys and orthogonal design
*Computer Modelling & New Technologies 2014 **18**(9) 510-513*

This paper attempts to use a simple and practical method based on the finite element theory and orthogonal design for the optimal structural parameters of machine tool bed. Dynamics finite element model of MCH63 machine tool bed has been established with the finite element analytic technique in order to improve the dynamic characteristics. The thickness of the ribs are looked as the design variables, the orthogonal design is used to optimize the natural frequency and the total mass. The simulation result indicates that the method is simple, effective and feasible.

Keywords: machine tool bed, optimal design, modal analysis, orthogonal design

Yinhui Long, Zilong Liao, Zhongxiao Guo, Yifan Song Evaluation of development patterns based on water resources carrying capacity calculation model in pastoral area
*Computer Modelling & New Technologies 2014 **18**(9) 514-517*

As the natural barrier, the health situation of grassland is relate to the change of social and economic development. According the natural environment of low rainfall and high evaporation, a comprehensive calculation model is established for evaluating the development patterns based on water resources carrying capacity. In present year, the social and grassland carrying capacity of BAILINGM is in the overload situation and there is a wide variance among different administrative region. By designing five different development patterns, we can find that carrying out the most strict water resources management policy can solve the overload problem of grassland carrying capacity effectively and the B-2 is the best development pattern. This paper show that the calculation results can provide reliable information and guide for macroscopic water resources management in large watershed and administrative region.

Keywords: development pattern, water resources, carrying capacity model, pastoral area

Jianguo Wang, Cong Cong, Xuhua Qing, Xiao Han The application of saturation memetic algorithm in economic load dispatch
*Computer Modelling & New Technologies 2014 **18**(9) 518-521*

To solve a nonlinear constrained economic load dispatch (ELD) problem in which the minimized generation cost is taken as objective and valve point effect of thermal units and operation constraints of power grid are taken into account. A saturation memetic algorithm was presented. To avoid the blindness in the search by the proposed using saturation increase efficiency and using adaptive penalty function to deal constraints. Calculation results of 3-machine system verify the effectiveness of the proposed algorithm. Compared with the other existing techniques, the proposed algorithm has been found to perform better. This method seems to be a promising alternative approach for solving the ELD problems in practical power system.

Keywords: memetic algorithm, economic load dispatch, valve point, saturation

Fenglei Han, Ankang Hu, Yachong Liu, Chunhui Wang Modelling and hull vibration calculation of very large container ship
*Computer Modelling & New Technologies 2014 **18**(9) 522-527*

Hull vibration is inevitable in large vessel's operation. Too large hull vibration will not only cause damage to the hull structure, but also will affect the normal use of marine equipment and crew at work and life. Therefore, to predict the performance of the ship's structure vibration in ship design stage is essential, which can guide reasonable solutions and necessary damping measures. The vibration calculation report includes: Ship's free vibration calculation and forced vibration response calculations caused by vibration source on board the ship. The former is mainly to obtain the free vibration characteristics of the whole ship structure, to find hull's 1~3 order of vibration modes, which are most likely to occur and the corresponding frequency values. The latter is mainly get the vibration response of hull ship's forced vibration caused by excitation source, to avoid low-frequency vibration's affect to the work and life of the crew. Its response should meet the relevant standards and requirements.

Keywords: modelling, free vibration, forced vibration, container ship, FEM

Xiaobin Ning, Ning Li, Junping Jiang Experiment of energy recovery efficiency and simulation research on EV's regenerative braking system
*Computer Modelling & New Technologies 2014 **18**(9) 528-533*

The proposed technique ECPS aims to design and implement a constant pressure hydraulic regenerative braking system with flywheel that can operate the problems of short driving range and improve efficiency of braking energy recovery about electrical vehicle. We established ECPS after comparing several hydraulic regenerative braking schemes .Then we explored the correlation between displacement of pump/motor and efficiency of energy recovery by undertook experiment on using variable displacement of pump/motor. After that, we investigated braking and ECE-15 driving condition simulation based on AMESim, evaluated correlation between displacement of pump/motor and efficiency of energy recovery. The results indicate that the driving range of electrical vehicle has been increased by 25% and the service life of battery was prolonged because depth of discharge was decreased.

Keywords: recovery efficiency, regenerative braking, experiment

Chunfu Zhang, Jianguo Miao, Song Tang Thrust line measurement of multi-nozzle solid rocket engine
*Computer Modelling & New Technologies 2014 **18**(9) 534-537*

Static thrust line is an important measuring parameter of solid rocket engine. Factors that affect the misalignment and deflection of thrust line of multi-nozzle solid rocket engine were analysed in static condition, and a measuring method by Laser Tracker for thrust line was researched. A calculation model of the thrust line was established based on simplified force spiral rule, which synthesizes the space force systems. The simulation results of misalignment and deflection indicate that the method has high measurement accuracy.

Keywords: solid rocket motor, static thrust line, multi-nozzle, laser tracker

Wuwei Feng Vibration characteristic researching of capillary of ultrasonic wire bonding
*Computer Modelling & New Technologies 2014 **18**(9) 538-542*

Wire bonding, a process of the connection between a semiconductor chip and a lead frame by a thin metal wire, is one of the important processes of electronic packaging. Vibration characteristic and friction behaviour of capillary of Microelectronic Packaging ultrasonic bonding system are studied. A dynamic contact model of capillary was built by finite element method to gain bonding mechanism of wire bonding system. Vibration response and contact friction property of the capillary are calculated. The effects of the loading frequency and static compressive force to the vibration response and contact friction stress were obtained, and the relationships between the loading frequency, the static compressive force and the vibration response were given. The researching conclusions can be used as references for having a better understanding of the bonding mechanism and the study of fault diagnose techniques for bonding process.

Keywords: ultrasonic wire bonding, contact analysis, finite element, capillary

Ping Li, Yongchi Li, Shiwei Duan, Mingtao Liu, Ruiyuan Huang Micro-macroscopic statistical description on

damage evolution of concrete
*Computer Modelling & New Technologies 2014 **18**(9) 543-549*

In this paper, we present two types of damage model, coupled pressure-shear damage model and tensile damage model, according to the different stress state and different damage mechanism in concrete. Combining microscopic and macroscopic mechanics, we derive the damage evolution equation and obtain relevant material parameters by fitting the test data of one-dimensional compression and tension test. In order to verify the proposed damage model, we carry out numerical simulation on wave propagation problems caused by the explosive charge in concrete columns. The simulation results are consistent with experimental results, which show successfully the evolution of damage in explosion process.

Keywords: concrete, coupled pressure-shear damage, tensile damage, damage evolution

Jianying Liu, Qiaoxin Zhang, Lixia Deng Numerical simulation for mine rescue capsule gas explosion dynamic response
Computer Modelling & New Technologies 2014 18(9) 550-554

In order to ensure the research and development production of coal mine "KJYF - 96/8 portable hardware capsule" in the coal mine use can effectively resist gas explosion shock pressure, structural strength meet the safety requirements. The finite element model is set up, to take specific gas coal dust explosion triangle under the action of shock waves, the stress, strain and displacement response. The results show that the peak of 0.6 MPa and 1.2 MPa, 300ms triangle shock wave, the capsule basic no plastic deformation, overall no failure parts, ships were able to be whole, the structure safety. Depending on the overall displacement nephogram view of displacement nephogram and components parts after shock wave action, maintain the overall coordination deformation, minimal relative displacement, sealing well.

Keywords: capsule, gas explosion, shock wave, dynamic response, numerical simulation

Youneng Zhang, Dawen Wang, Li Ding A disturbance detection interferometric fibre optic sensor system
*Computer Modelling & New Technologies 2014 **18**(9) 555-559*

This study proposed a Mach-Zehnder interferometer (M-Z)-based disturbance detection interferometric fibre optic sensor system. In this system, two single-mode optical fibres passing through the central tube of cable were used as the signal arm and reference arm of the M-Z interferometer. In the coherent length range of laser light source, as long as encountering suitable optical path difference, can the photosynthesis generate a series of alternately dark and bright interference fringes. Reference arm is dependent from the external signals, while the signal arm is subject to the disturbance of external signals. Affected by the disturbance, signal interference fringes significantly generated fluctuations with the variation of the optical distance between the two arms. The interference signals were then collected and stored through digital storage oscilloscope. By detecting the variations of the interference fringes from stability to instability, and from regularity to irregularity, it had access to judging whether or not there was external disturbance invasion. Experimental results suggested that the sensor system was very sensitive. In case of slight touch on the cable, the waveform will generate variations.

Keywords: interferometer, optical fibre senescing, disturbance, optical path difference

Jingbin Hao, Haifeng Yang, Zhongkai Li Numerical simulation of aerodynamic characteristics of windmill based on computer graphics display technology
*Computer Modelling & New Technologies 2014 **18**(9) 560-564*

This paper proposed research on wind computation fluid dynamics(CFD) analysis and visualization in scientific computing facing to OSG in order to obtain numerical result of aerodynamic characteristics stimulated by windmill and to display vivid and direct 3-D effect for easy observation experiment. This paper applied CFX module in ANSYS Workbench software to conduct CFD analysis and calculation of windmill through computer graphic display technology based on CFD principle and realize numerical stimulation of aerodynamic characteristics of windmill. Then the research on visualization in scientific computing of aerodynamic characteristics of windmill was realized combining with OSG software that face to virtual display technology

Keywords: computation fluid dynamics, computer graphics display technology, numerical stimulation, 3-D geometric model