



COMPUTER MODELLING
AND
NEW TECHNOLOGIES

2014
VOLUME 18 NO 5

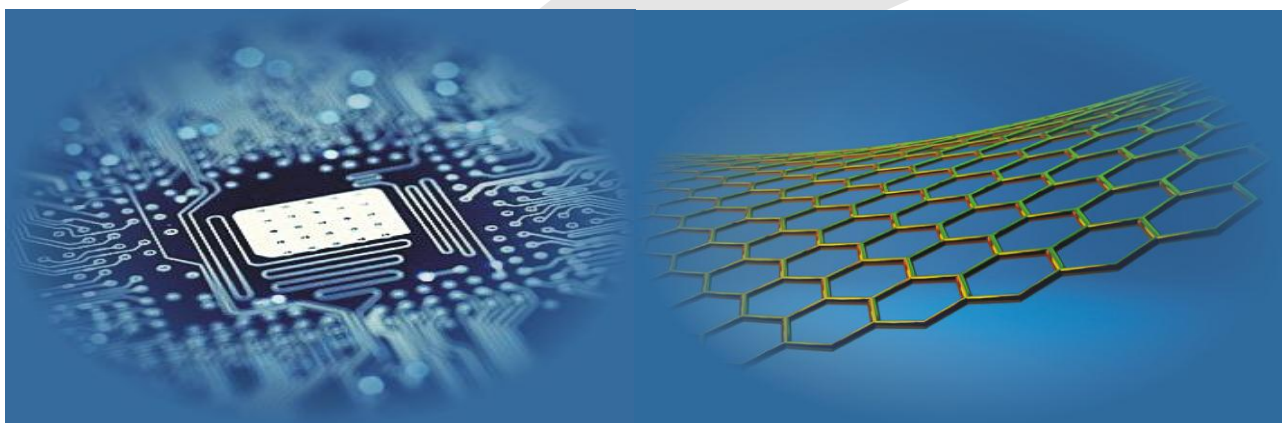
ISSN 1407-5806 ISSN 1407-5814 on-line

Transport and Telecommunication Institute
and
Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2014 Volume 18 No 5

ISSN 1407-5806, ISSN 1407-5814 (*On-line: www.tsi.lv*)



Riga – 2014

EDITORIAL BOARD

Prof. Igor Kabashkin	Chairman of the Board , <i>Transport & Telecommunication Institute, Latvia</i>
Prof. Yuri Shunin	Editor-in-Chief , <i>Information Systems Management Institute, Latvia</i>
Prof. Adolfas Baublys	<i>Vilnius Gediminas Technical University, Lithuania</i>
Dr. Brent Bowen	<i>Embry-Riddle Aeronautical University, United States of America</i>
Prof. Olgierd Dumbrajs	<i>University of Latvia, Solid State Physics Institute, Latvia</i>
Prof. Sergey Maksimenko	<i>Institute for Nuclear Problem, Belarus State University, Belarus</i>
Prof. Vladimir Litovchenko	<i>V. Lashkaryov Institute of Semiconductor Physics of National Academy of Science of Ukraine, Ukraine</i>
Prof. Pavel D'yachkov	<i>Kurnakov Institute for General and Inorganic Chemistry, Russian Academy of Sciences, Russian Federation</i>
Prof. Stefano Bellucci	<i>Frascati National Laboratories – National Institute of Nuclear Physics, Italy</i>
Prof. Arnold Kiv	<i>Ben-Gurion University of the Negev, Israel</i>
Prof. Alytis Gruodis	<i>Vilnius University, Lithuania</i>
Prof. Michael Schenk	<i>Fraunhofer Institute for Factory Operation and Automation IFF, Germany</i>
Prof. Dietmar Fink	<i>University of Mexico, United Mexican States</i>
Prof. Ravil Muhamedyev	<i>International IT University, Kazakhstan</i>
Prof. Kurt Schwartz	<i>Gesellschaft für Schwerionenforschung mbH, Darmstadt, Germany</i>
Prof. Eva Rysiakiewicz-Pasek	<i>Institute of Physics, Wroclaw University of Technology, Poland</i>
Contributing Editor	Prof. Victor Gopeyenko, <i>Information Systems Management Institute, Latvia</i>
Literary Editor	Prof. Tamara Lobanova-Shunina, <i>Riga Technical University, Latvia</i>
Technical Editor , secretary of Editorial Board	MSc Comp Nataly Burluckaya, <i>Information Systems Management Institute, Latvia</i>

Journal topics:	Host Organization	Supporting Organizations
mathematical and computer modelling computer and information technologies natural and engineering sciences operation research and decision making nanoscience and nanotechnologies innovative education	Transport and Telecommunication Institute	Latvian Transport Development and Education Association Latvian Academy of Sciences Latvian Operations Research Society
Articles should be submitted in English . All articles are reviewed.		

EDITORIAL CORRESPONDENCE	COMPUTER MODELLING AND NEW TECHNOLOGIES, 2014, Vol. 18, No.5 ISSN 1407-5806, ISSN 1407-5814 (on-line: www.tsi.lv)
Transport and Telecommunication Institute	Scientific and research journal The journal is being published since 1996
1 Lomonosova, Bld 4 , LV-1019, Riga, Latvia Phone: (+371) 67100594 Fax: (+371) 67100535 E-mail: cm&nt@tsi.lv www.tsi.lv	The papers published in Journal 'Computer Modelling and New Technologies' are included in: INSPEC (since 2010) , www.theiet.org/resources/inspec/ VINITI (since 2011) , http://www2.viniti.ru/ CAS Database http://www.cas.org/ El Compindex



Content

Editors' Remarks		5
Mathematical and Computer Modelling		
Jun Liu, Xing He, Qingyou Liu, Jiang Naibin, Huang Chen	Vibration-modal analysis model for multi-span pipeline with different support conditions	7
Ni-qin Jing, Lin-na Wang	The research of electromotor control based on optimized RBF neural network	14
Chang-liang Liu, Zeng-hui Ma, Ping-an Kai	A universal tuning method of large dead-time system	19
Wenjing Li	Using grey-weighted Markov chain model to predict the quantum of highway passenger transport	24
Hongyang Zhang	Research on modelling of intake tower in three-dimension CAD software and simulation analysis in FE software	29
Wenliang Yin, Maoqing Xiang	Research on the operation modes of hydropower station based on complementary characteristics	36
Hai Liu, Jing Chen, Jian-yi Kang, Xiao-xia Li, Ivan Azhari	Finite element analysis of the dynamic response of the cardiovascular system to the blunt ballistic impact	44
Nanjian Zhuang, Jinwu Xiang, Zhangping Luo, Yiru Ren	Calculation of helicopter maneuverability in forward flight based on energy method	50
Shu Jie Liu, Ya Wei Hu, Chao Li, Hong Chao Zhang	Residual life prediction under condition monitoring	55
Qing Ma, Hua Yang, Chaogang Zhang, Zhaohui Peng	Effects of global warming for building energy demand in China	61
Weijun Cheng	On the error rate analysis of dual-hop relaying over composite fading channels using mixture gamma distribution	66
Computer and Information Technologies		
Gaochao Xu, Peng Liu, Xiaodong Fu, Yunmeng Dong, Jia Zhao, Yan Ding	A novel task deployment approach based on graph theory for power saving	73
Juncheng Li	Image enlargement based on the hyperbolic coons interpolation surface	79
Yiping Shen, Shuxiao Li, Chengfei Zhu, Hongxing Chang	A fast top-down visual attention method to accelerate template matching	86
Weidong Tang, Jinzhao Wu, Meiling Liu	Step semantics and action refinement in event structures	94
Xianfeng Yang, Yan Wang	A research into static traffic routing and resource optimization algorithm based on genetic and tabu search	102
Xianmin Wei, Hong Lu, Hong Feng	ABMP: adaptive bitmap protocol within TDMA for mobile underwater sensor networks	111
Hong Li, ZongZhe Wang	Design of gray PID controller for DC servo motor	118
Bingqian Chen	A new automatic registration method for InSAR image based on multi-step strategy	124
Lirong Qiu	An approach for reference model implementation by predicting all possible output of design	131
Rong-chun Wu, Feng-li Zhang, Jin-bang Zhang, Qian He	Application of fuzzy comprehensive evaluation in weapon equipment systems	138

Operation research and decision making		
Yingjun Zhang, Xuefeng Yang	Expert system based on fuzzy rules for maritime search and rescue	143
Jun Luo, Sijing Cai, Yanhui Wang	Research on dynamic risk identification model of shield tunnelling based on reason model	149
Yanming Ye, Yueshen Xu, Zhilin Feng	Tag-based process recommendation for social business process modelling	155
Jin Wang, Weidong Zhu	The impact of capital structure on corporate performance based on panel threshold model	162
Hong Zhang	Short-term prediction of wind power based on self-adaptive niche particle swarm optimization	168
C M Xue, Y Yang, T Yang, T T Zeng	Person-organization fit evaluation and process optimization based on the matching theory	174
Jinling Li, Haixiang Guo, Yan Chen, Deyun Wang, Kejun Zhu	An artificial fish swarm algorithm for solving a bi-objective capacitated vehicle routing problem	181
Lijuan Wang, Song Jin, Tianwei Zhang, Peter Chung	Analysis of price rate models for household water consumption in urban China	191
HaiLei Zou, Cheng Wang	Calculation of China's environmental efficiency based on the SBM model with undesirable outputs	196
Zhaoxia Si, Shaoliang Zhang, Ningli Chen	Correlation model analysis on the land price fluctuations in Beijing and Tianjin City in China	201
Haiyan Yi, Dianjun Fang	Dynamic evaluation and simulation of variant-driven complexity costs in multi-echelon automotive supply chains	208
Chenghu Yang, Lanying Liu, Lei Zhang	Optimal acquisition and pricing policies for remanufacturing systems with initial investment	220
Xing Yu	Continuous-time optimal portfolio model with mean-reverting process	226
Nannan Duan, Fuyuan Xu, Ming Ni	Evolutionary game analysis of enterprises' technological innovation strategies	230
Nature Phenomena and Innovative Engineering		
Lixiong Gong, Xiangsheng Kong, Yong Liu, Min Huang	Subpixel edge extraction of part ant colony optimization-based and dimensional measurement	240
Hong Li	Application of PID-type iterative learning control for DC motor	247
Xianliang Yang, Lianlian Jia, Songling Wang, Jiangjiang Wang	Based on pressure gradient model to determine leakage point in heating pipe network	252
Jian-Ning Han, Peng Yang, Lu Zhang	External locating of moving targets for 3D IMRT using parallax method	257
Chun Li, Hong Nie, Jinbao Chen	Motion analysis and simulation of a 12-Tetrahedral Walker Robot	263
W Huang, D Y Liu, H F Jiang, F Y Liu	Numerical and analytical solution of stresses on a box-type lining structure under the effect of ground fracture	269
Zhenzhen Jia, Feng Tao	Numerical simulation on methane explosion propagation in a one-dimensional straight duct with porous metal materials	275
Bin Dong, Guojie Yang, Tie Liang	Research and realization of handheld radio direction finding communication system	283
Jiong Zhang, Shaofei Chen, Lihua Liu	Research on commutation torque ripple suppression strategy of BLDCM based on iterative learning	288
Guoping Shi, Jun Liang	The gradual learning static load modelling method based on real-time fault recorder data	297
Ying Kong, Xiao guang Chu	The scroll flow and torque prediction with the wavelet neural network optimized by PSO and BP	303
Xiaohui Liu, Xiaoping Zhao, Jianfeng Liu	A study on the acoustic emission characteristics of the coal rock on different bedding direction	308
Authors' Index		314
Cumulative Index		315

Editors' Remarks

A Legend of Truth

by Rudyard Kipling

Once on a time, the ancient legends tell,
 Truth, rising from the bottom of her well,
 Looked on the world, but, hearing how it lied,
 Returned to her seclusion horrified.
 There she abode, so conscious of her worth,
 Not even Pilate's Question called her forth,
 Nor Galileo, kneeling to deny
 The Laws that hold our Planet 'neath the sky.
 Meantime, her kindlier sister, whom men call
 Fiction, did all her work and more than all,
 With so much zeal, devotion, tact, and care,
 That no one noticed Truth was elsewhere.

Then came a War when, bombed and gassed and mined,
 Truth rose once more, perforce, to meet mankind,
 And through the dust and glare and wreck of things,
 Beheld a phantom on unbalanced wings,
 Reeling and groping, dazed, dishevelled, dumb,
 But semaphoring direr deeds to come.

Truth hailed and bade her stand; the quavering shade
 Clung to her knees and babbled, "Sister, aid!
 I am - I was - thy Deputy, and men
 Besought me for my useful tongue or pen
 To gloss their gentle deeds, and I complied,

And they, and thy demands, were satisfied.
 But this - "she pointed o'er the blistered plain,
 Where men as Gods and devils wrought amain--
 "This is beyond me! Take thy work again."

Tablets and pen transferred, she fled afar,
 And Truth assumed the record of the War...
 She saw, she heard, she read, she tried to tell
 Facts beyond precedent and parallel--
 Unfit to hint or breathe, much less to write,
 But happening every minute, day and night.
 She called for proof. It came. The dossiers grew.
 She marked them, first, "Return. This can't be true."
 Then, underneath the cold official word:
 "This is not really half of what occurred."

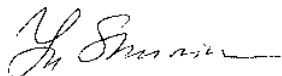
She faced herself at last, the story runs,
 And telegraphed her sister: "Come at once.
 Facts out of hand. Unable overtake
 Without your aid. Come back for Truth's own sake!
 Co-equal rank and powers if you agree.
 They need us both, but you far more than me!"

Rudyard Kipling (1809-1849) *


This 18th volume No.5 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".



Vibration-modal analysis model for multi-span pipeline with different support conditions

Jun Liu^{1, 2}, Xing He¹, Qingyou Liu^{2*}, Jiang Naibin³, Huang Chen¹

¹*School of Civil Engineering and Architecture, Southwest Petroleum University, Chengdu 610500, China*

²*Modern Design and Simulation Lab for Oil&Gas Equipment's, Southwest Petroleum University, Chengdu 610500, China*

³*Science and Technology on Reactor System Design Technology Laboratory, Nuclear Power Institute of China, Chengdu 610041, China*

Received 21 May 2014, www.tsi.lv

Abstract

Vibration characteristics analysis is important in the design of multi-span pipeline with different support conditions. In order to analyze the natural frequency and the vibration modal of the multi-span pipeline, a matrix transfer method is proposed in this paper. With the multi-span pipeline divided into single-span pipes, the transmission formulas for the deflection, angle, bending and shear between two adjacent spans are deduced, in combination with the Krylov function solution of the free vibration equation for the single-span pipe, and the constraint condition between the two adjacent spans of the multi-span pipeline. According to the boundary conditions on the starting and ending spans, the natural frequency equation and the vibration modal function between two adjacent spans of the multi-span pipeline are presented. The FORTRAN program based on the above principle is written, and the natural vibration frequencies and the vibration modals of two typical multi-span pipelines are investigated and compared with the results from ABAQUS. It is shown that the model presented in this paper is efficient in the analysis of multi-span pipeline and has the advantages of high computational efficiency and convenience for engineering practice application.

Keywords: Multi-span pipeline, Vibration equation, Natural frequency, Vibration modal

1 Introduction

With widely used in petroleum engineering, chemical engineering and nuclear power, Multi-span flow pipeline is one of the most common industrial equipment's. During the design of the Multi-span flow pipeline, in order to prevent the production of "instability" and "resonance" phenomenon, much attention should be paid to its dynamic characteristics such as natural frequency and vibration modal. In order to do it, various approaches have been used ranging from numerical methods such as finite element method [1, 2] to analytical models [3]. The former, although has high calculation accuracy, is inconvenience to be used. The latter with appropriate accuracy and relatively low computational cost is convenient to be used in engineering.

For the analytical models, traditionally, three-bending-moment model [4], which is based on the continuous condition of the angle and bending moment on the support linked to two adjacent spans, is usually used to establish frequency equation and solved it by numerical analysis method. This model shows good adaptability to low-order frequencies but for the calculation of high-order frequency, the computational precise will decrease due to the increase of the iteration number. Zhang et al and others [5-10] successively presented the frequency equations of multi-span pipeline with flexible and rigid supports. The influence of support

to vibration modal had been also investigated by them. These models also appear capability in the determination of low-order frequencies of multi-span pipeline. However, the derivation processes of vibration mode function and frequency function in them are not available and high-order frequency solution with higher precise still remains unsatisfied. Therefore, it is clearly desirable to develop an effective analytical model with better precise for high-order vibration and more convenient for practical application.

The aim of this paper is to provide an efficient method of vibration characteristics analysis to the design of multi-span pipeline with different spans and different supports. The continuity condition of multi-span pipeline is used to derive the vibration modal functions of the multi-span flow pipeline and by solving them with iteration technique, high-order natural frequency and vibration modal with higher precise can be obtained.

2. Free vibration equation for multi-span flow pipeline

2.1 SINGLE-SPAN FLOW PIPELINE

Based on vibration analysis theory for single-span beam, vibration equation for multi-span flow pipeline can be derived [11-13]. Therefore, according to the d'Alembert principle, the differential equation of free vibration for

* *Corresponding author* e-mail: Liuqy66@aliyun.com

single-span beam with material homogeneous and cross section uniform (see figure 1) can be written as follows

$$EI \frac{\partial^4 y(x,t)}{\partial x^4} + \rho A \frac{\partial^2 y}{\partial t^2} = 0, \quad (1)$$

where EI is bending stiffness, ρ is material density, A is the area of cross section.

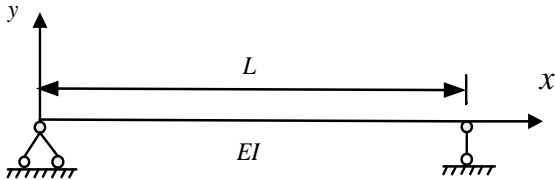


FIGURE 1 Single-span continuous beam with two ends simply supported

Formula (1) is a fourth-order differential equation and its general solution is written as follows, $y(x,t) = [AS(kx) + BT(kx) + CU(kx) + DV(kx)] \sin pt$, where A, B, C, D are constants, $\sin pt$ is the variable that considers the change of displacement with time and has no effect on the natural frequency as well as the vibration mode of the multi-span flow pipeline, $S(kx), T(kx), U(kx)$ and $V(kx)$ are the functions of hyperbolic functions and trigonometric functions, and the expansion of them can be written as follows:

$$S(kx) = \frac{\text{ch}(kx) + \cos(kx)}{2}, \quad T(kx) = \frac{\text{sh}(kx) + \sin(kx)}{2},$$

$$U(kx) = \frac{\text{ch}(kx) - \cos(kx)}{2}, \quad V(kx) = \frac{\text{sh}(kx) - \sin(kx)}{2}, \text{ where}$$

k is the function of bending stiffness, material density, area of cross section and natural circular frequency and:

$$k^4 = \frac{EI}{\rho A} \omega^2, \text{ where } \omega \text{ is the natural circular frequency (rad / s).}$$

2.2 MULTI-SPAN FLOW PIPELINE

Multi-span continuous flow pipeline is regarded as the structure that is composed of several single-span beams. Each of them has the same vibration mode as a single-span beam. Therefore, for the multi-span pipeline as shown in Figure 2, the general solution for the free vibration equation of the i^{th} span can be written as:

$$y_i(x,t) = [A_i S(kx) + B_i T(kx) + C_i U(kx) + D_i V(kx)] \sin pt. \quad (2)$$

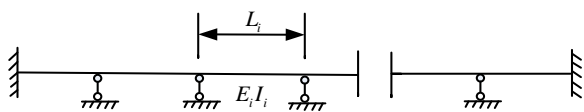


FIGURE 2 Multi-span continuous beams with fixed ends and rigid supports

Based on formula (2), the following expression can be used as the vibration modal function of the span:

$$\varphi_i(x) = y_i(x) = A_i S(kx) + B_i T(kx) + C_i U(kx) + D_i V(kx). \quad (3)$$

It appears that the vibration modal function is composed of the displacements of each cross section at some time. Thus, the displacement, rotation, bending moment and shear force of each cross section of the span can be expressed as formulas (4)-(7):

$$y_i(x) = A_i S(kx) + B_i T(kx) + C_i U(kx) + D_i V(kx), \quad (4)$$

$$\theta_i(x) = \frac{\partial y_i(x)}{\partial x} = k [A_i V(kx) + B_i S(kx) + C_i T(kx) + D_i U(kx)], \quad (5)$$

$$M_i(x) = EI \frac{\partial^2 y_i(x)}{\partial x^2} = EI k^2 [A_i U(kx) + B_i V(kx) + C_i S(kx) + D_i T(kx)], \quad (6)$$

$$Q_i(x) = EI \frac{\partial^3 y_i(x)}{\partial x^3} = EI k^3 [A_i T(kx) + B_i U(kx) + C_i V(kx) + D_i S(kx)]. \quad (7)$$

The force analysis diagram for the i^{th} span is shown as Figure 3, in which y_i, θ_i, M_i and Q_i respectively denote displacement, rotation, bending moment and shear force, superscript L, R respectively denote left end and right end.

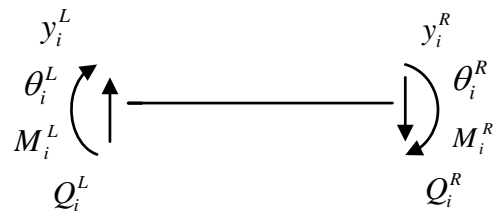


FIGURE 3 Force analysis diagram for i^{th} span

At the left end ($x = 0$), formulas (4)-(7) can be simplified as follows:

$$\begin{cases} y_{iL} = A_i \\ \theta_{iL} = kB_i \\ M_{iL} = k^2 EI C_i \\ Q_{iL} = k^3 EI D_i \end{cases}. \quad (8)$$

According to equation (8), the four constants of i^{th} span A_i, B_i, C_i, D_i can be determined as follows:

$$A_i = y_{iL}, \quad B_i = \frac{\theta_{iL}}{k}, \quad C_i = \frac{M_{iL}}{k^2 EI}, \quad D_i = \frac{Q_{iL}}{k^3 EI}.$$

With the four constants substituted into formula (4)-(7), the displacement, rotation, bending moment and shear force equation of the right end ($x = L_i$) of the i^{th} span can be given by:

$$\begin{cases} y_{iR} = S_i y_{iL} + \frac{T_i}{k} \theta_{iL} + \frac{U_i}{k^2 EI} M_{iL} + \frac{V_i}{k^3 EI} Q_{iL} \\ \theta_{iR} = kV_i y_{iL} + S_i \theta_{iL} + \frac{T_i}{kEI} M_{iL} + \frac{U_i}{k^2 EI} Q_{iL} \\ M_{iR} = E_i I k^2 U_i y_{iL} + E_i I k V_i \theta_{iL} + S_i M_{iL} + \frac{T_i}{k} Q_{iL} \\ Q_{iR} = E_i I k^3 T_i y_{iL} + E_i I k^2 U_i \theta_{iL} + kV_i M_{iL} + S_i Q_{iL} \end{cases} \quad (9)$$

The equation (9) can be written as matrix form

$$\begin{bmatrix} y_{iR} \\ \theta_{iR} \\ M_{iR} \\ Q_{iR} \end{bmatrix} = [C]_{4 \times 4} \begin{bmatrix} y_{iL} \\ \theta_{iL} \\ M_{iL} \\ Q_{iL} \end{bmatrix} \quad (10)$$

The items in the matrix $[C]_{4 \times 4}$ are: $C_{11} = S_i$,

$$C_{12} = \frac{T_i}{k}, \quad C_{13} = \frac{U_i}{k^2 EI}, \quad C_{14} = \frac{V_i}{k^3 EI}, \quad C_{21} = kV_i,$$

$$C_{22} = S_i, \quad C_{23} = \frac{T_i}{kEI}, \quad C_{24} = \frac{U_i}{k^2 EI}, \quad C_{31} = E_i I k^2 U_i,$$

$$C_{32} = E_i I k V_i, \quad C_{33} = S_i, \quad C_{34} = \frac{T_i}{k}, \quad C_{41} = E_i I k^3 T_i,$$

$$C_{42} = E_i I k^2 U_i, \quad C_{43} = kV_i, \quad C_{44} = S_i.$$

$$\begin{bmatrix} y_{2R} \\ \theta_{2R} \\ M_{2R} \\ Q_{2R} \end{bmatrix} = \begin{bmatrix} C_{13}^2 & C_{14}^2 \\ C_{23}^2 & C_{24}^2 \\ C_{33}^2 & C_{34}^2 \\ C_{43}^2 & C_{44}^2 \end{bmatrix} \begin{bmatrix} \theta_{2L} \\ Q_{2L} \end{bmatrix} = [C]_{4 \times 4}^2 \begin{bmatrix} 0 & 0 \\ -C_{23}^1 \frac{C_{13}^1}{C_{14}^1} + C_{24}^1 & 0 \\ -C_{33}^1 \frac{C_{13}^1}{C_{14}^1} + C_{34}^1 & 0 \\ -C_{43}^1 \frac{C_{13}^1}{C_{14}^1} + C_{44}^1 & 1 \end{bmatrix} \begin{bmatrix} Q_{1L} \\ F_1 \end{bmatrix} \quad (12)$$

Using the similar derivation process as formula (12), the displacement, rotation, bending moment and shear force in the right end of the i^{th} span can be obtained as:

$$\begin{bmatrix} y_{iR} \\ \theta_{iR} \\ M_{iR} \\ Q_{iR} \end{bmatrix} = \begin{bmatrix} C_{13}^i & C_{14}^i \\ C_{23}^i & C_{24}^i \\ C_{33}^i & C_{34}^i \\ C_{43}^i & C_{44}^i \end{bmatrix} \begin{bmatrix} \theta_{2L} \\ Q_{2L} \end{bmatrix} = [C]_{4 \times 4}^i \begin{bmatrix} 0 & 0 \\ -C_{23}^{i-1} \frac{C_{13}^{i-1}}{C_{14}^{i-1}} + C_{24}^{i-1} & 0 \\ -C_{33}^{i-1} \frac{C_{13}^{i-1}}{C_{14}^{i-1}} + C_{34}^{i-1} & 0 \\ -C_{43}^{i-1} \frac{C_{13}^{i-1}}{C_{14}^{i-1}} + C_{44}^{i-1} & 1 \end{bmatrix} \begin{bmatrix} Q_{1L} \\ F_i \end{bmatrix} \quad (13)$$

The equation (10) also set up the force transitive relation between right end and left end of the span. This means that, as long as the internal force in the left end is determined, the internal force in the right end can also be deduced by equation (10). For example, in the first span shown in figure 2, since the displacement and bending moment of the left end with a fixed bearing are zero, the equation (10) can be simplified as:

$$\begin{bmatrix} y_{1R} \\ \theta_{1R} \\ M_{1R} \\ Q_{1R} \end{bmatrix} = [C]_{4 \times 4} \begin{bmatrix} y_{1L} \\ \theta_{1L} \\ M_{1L} \\ Q_{1L} \end{bmatrix} = \begin{bmatrix} C_{13}^1 & C_{14}^1 \\ C_{23}^1 & C_{24}^1 \\ C_{33}^1 & C_{34}^1 \\ C_{43}^1 & C_{44}^1 \end{bmatrix} \begin{bmatrix} M_{1L} \\ Q_{1L} \end{bmatrix} \quad (11)$$

For the second span, considering the supporting condition on the left end of the pipeline, we have:

$$\begin{bmatrix} y_{2L} \\ \theta_{2L} \\ M_{2L} \\ Q_{2L} \end{bmatrix} = \begin{bmatrix} y_{1R} \\ \theta_{1R} \\ M_{1R} \\ Q_{1R} + F_1 \end{bmatrix}, \text{ where } F_1 \text{ is the bearing shear in}$$

the left end of the 1th span.

Using the above relation and formula (10), the force transitive relation between the right end of 2th span and the left end of the 1th span can be derived as follows:

If the right end of the pipeline is supported by hinge support, then $y_{iR} = M_{iR} = 0$. According to formula (13),

we have $\begin{bmatrix} C_{13}^i & C_{14}^i \\ C_{33}^i & C_{34}^i \end{bmatrix} = 0$. If the right end is supported by

fixed bearing, then $y_{iR} = \theta_{iR} = 0$, we have

$$\begin{bmatrix} C_{13}^i & C_{14}^i \\ C_{23}^i & C_{24}^i \end{bmatrix} = 0.$$

Based on the condition that the determinant value of the matrix C is zero, a test-value method can be used to determine the natural circular frequency of the multi-span continuous pipeline. In the method, different k values are substituted into the determinant of matrix C. if the determinant value corresponding to a certain k is zero, the natural circular frequency is obtained using the formula

$$k^4 = \frac{EI}{\rho A} \omega^2.$$

Sorting these obtained k values from small to large order, the natural circular frequencies from low order to high order can be obtained.

If the left end of the first span is supported by hinge support and elastic supports are used on the other span

$$\begin{bmatrix} y_{1R} \\ \theta_{1R} \\ M_{1R} \\ Q_{1R} \end{bmatrix} = [C]_{4 \times 4} \begin{bmatrix} y_{1L} \\ \theta_{1L} \\ M_{1L} \\ Q_{1L} \end{bmatrix} = \begin{bmatrix} C_{12}^1 & C_{14}^1 \\ C_{22}^1 & C_{24}^1 \\ C_{32}^1 & C_{34}^1 \\ C_{43}^1 - K_i C_{12}^1 & C_{44}^1 - K_i C_{14}^1 \end{bmatrix} \begin{bmatrix} M_{1L} \\ Q_{1L} \end{bmatrix}. \quad (15)$$

By means of the method used in deriving formula (13), the recursion relation of bearing internal force of this kind of pipeline supported by mixture supports can be obtained.

$$\begin{bmatrix} y_{iR} \\ \theta_{iR} \\ M_{iR} \\ Q_{iR} \end{bmatrix} = \begin{bmatrix} C_{12}^{i-1} & C_{14}^{i-1} \\ C_{22}^{i-1} & C_{24}^{i-1} \\ C_{32}^{i-1} & C_{34}^{i-1} \\ C_{43}^{i-1} - K_i C_{12}^{i-1} & C_{44}^{i-1} - K_i C_{14}^{i-1} \end{bmatrix} \begin{bmatrix} M_{1L} \\ Q_{1L} \end{bmatrix}. \quad (16)$$

The natural circular frequencies of the multi-span pipeline with rigid supports are the values that make the determinant value of matrix C zero. The determinants of the matrix C is depended on support condition of the pipeline. The determinants of matrix C for two kinds of supports are listed as follows:

The *Fixed* support at the right end $\begin{bmatrix} C_{12}^i & C_{14}^i \\ C_{32}^i & C_{34}^i \end{bmatrix} = 0$.

The *Elastic* support at the right end $\begin{bmatrix} C_{32}^i & C_{34}^i \\ C_{43}^i - K_i C_{12}^i & C_{44}^i - K_i C_{14}^i \end{bmatrix} = 0$.

The *Test-value* method is

ends (see Figure 4), formula (11) is changed to formula (14).

$$\begin{bmatrix} y_{1R} \\ \theta_{1R} \\ M_{1R} \\ Q_{1R} \end{bmatrix} = [C]_{4 \times 4} \begin{bmatrix} y_{1L} \\ \theta_{1L} \\ M_{1L} \\ Q_{1L} \end{bmatrix} = \begin{bmatrix} C_{12}^1 & C_{14}^1 \\ C_{22}^1 & C_{24}^1 \\ C_{32}^1 & C_{34}^1 \\ C_{42}^1 & C_{44}^1 \end{bmatrix} \begin{bmatrix} \theta_{1L} \\ Q_{1L} \end{bmatrix}. \quad (14)$$

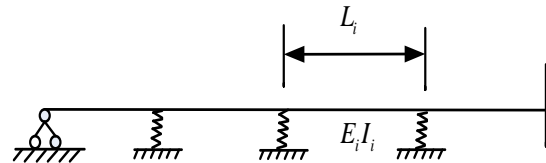


FIGURE 4 Multi-span pipeline with the first span supported by hinge and other spans elastically supported

Since the middle bearings are supported by elastic supports, therefore $Q_{iR} = Q_{iL} + K_i y_i$, where K_i is the stiffness coefficient of the first elastic support (KN/m). Thus, in this case, the formula (13) can be changed to the following:

also used to determine the natural circular frequencies of the multi-span continuous pipeline.

3 Modal function for multi-span continuous flow pipeline

Since each span of the multi-span pipeline has its vibration modal function, therefore the modal function analysis on the whole pipeline structure is to determine the four constants of the vibration modal function of each span. It is noted that the multi-span pipeline is continuous. Thus, the four constants of the vibration modal function for a span can be obtained by introducing the boundary condition and constraint condition between the span and its adjacent span into the modal function.

The modal function for the first span can be written as:

$$y_1(x) = A_1 S(kx) + B_1 T(kx) + C_1 U(kx) + D_1 V(kx). \quad (17)$$

It is assumed that the left end of the first span is supported by fixed bearing (see figure 2). Then the boundary condition for the left end of the first span can be described by the following equations

$$\begin{cases} A_1 = 0 \\ B_1 = 0 \\ C_1 U_1(kL_1) + D_1 V_1(kL_1) = 0 \end{cases} \quad (18)$$

If the support between the i^{th} span and $(i+1)^{\text{th}}$ span is hinge bearing, the constraint condition between the two spans can be expressed as follows

$$\begin{cases} A_i = 0 \\ B_{i+1} = B_i S_i(kL_i) + C_i T_i(kL_i) + D_i U_i(kL_i) \\ C_{i+1} = B_i V_i(kL_i) + C_i S_i(kL_i) + D_i T_i(kL_i) \\ B_i T_i(kL_i) + C_i U_i(kL_i) + D_i V_i(kL_i) = 0 \end{cases} \quad (19)$$

If the right end of the pipeline is supported by fixed bearing, then the boundary condition can be written as

$$\begin{cases} A_n = 0 \\ B_n = B_{n-1} S_{n-1}(kL_{n-1}) + C_{n-1} T_{n-1}(kL_{n-1}) + D_{n-1} U_{n-1}(kL_{n-1}) \\ B_n S_n(kL_n) + C_n T_n(kL_n) + D_n U_n(kL_n) = 0 \\ B_n T_n(kL_n) + C_n U_n(kL_n) + D_n V_n(kL_n) = 0 \end{cases} \quad (20)$$

The modal computing can be accomplished by means of the recursive method. In the method, the four constants for the modal function of the first span can be determined by assuming $C_1=1$ and introducing it into the equation (18). Base on the calculation for the first span, the other four constants for the second span can be obtained by using the equation (19). The similar recursive process is repeated until the four constants for the modal function of the last span is computed by using the equation (20).

Based on the theory presented by the above sections, the computer code of the method is developed using the FORTRAN language and the validity of the method will be proved in the next section by using the finite elements software ABAQUS.

4 Example

In order to verify the computer code, computations were carried out for multi-span flow pipelines with different type of supports.

4.1 MULTI-SPAN CONTINUOUS OIL PIPELINE WITH RIGID SUPPORTS

A multi-span oil pipeline with two ends fixed and middle part supported by rigid bearing is shown in Figure 5. The geometry and physical parameters of the pipeline are listed in Table 1.

Liu Jun, He Xing, Liu Qinqyou, Naibin Jiang, Chen Huang

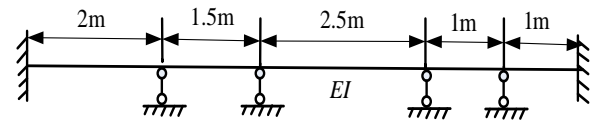


FIGURE 5 Unequal-span oil pipeline with rigid supports
TABLE 1 Geometry and physical parameters of the unequal-span pipeline with rigid supports

Elasticity modulus (GPa)	Pipeline density (kg/m ³)	Oil density (kg/m ³)	Pipe outside diameter (m)	Pipeline wall thickness (m)
200	7850	900	0.16	0.01

The computer code developed in this paper is used to analysis the vibration modal of the pipeline. A comparison between the calculation result of the first five order modal of the pipeline and that from ABAQUS (in the ABAQUS 1000 B21 elements are used) is carried out. The frequencies are listed in Table 2 and the vibration modals are shown in Figure 6.

TABLE 2 The first five natural frequencies of the unequal-span pipeline with rigid supports

Frequency order	Current method	ABAQUS
1	81.585	80.452
2	138.937	136.93
3	203.391	205.99
4	263.632	263.83
5	421.436	422.52
6	494.519	508.73
7	592.239	607.44
8	666.580	656.59
9	795.180	818.83
10	866.966	858.19
11	953.093	1009.5
12	1004.330	166.8
13	1292.535	1248.9
14	1308.915	1261.9
15	1408.384	1321.4

It is shown in Table 2 that the first five order frequencies respectively from the current method and ABAQUS are closed and the maximum error between them is less than 5%. In the Figure 6, it appears that the vibration modals from the model presented in this paper are essentially identical to those from ABAQUS. Therefore, the current model is efficient in the vibration analysis on unequal-span pipeline with rigid support.

4.2 MULTI-SPAN CONTINUOUS OIL PIPELINE WITH ELASTIC SUPPORTS

In this section, the vibration modal analysis on a multi-span oil pipeline with two ends fixed and middle part supported by elastic bearings is carried out. The geometry and physical parameters of the pipeline are the same as those listed in Table 1. The stiffness coefficients of all the four elastic bearings are 43 kN/m.

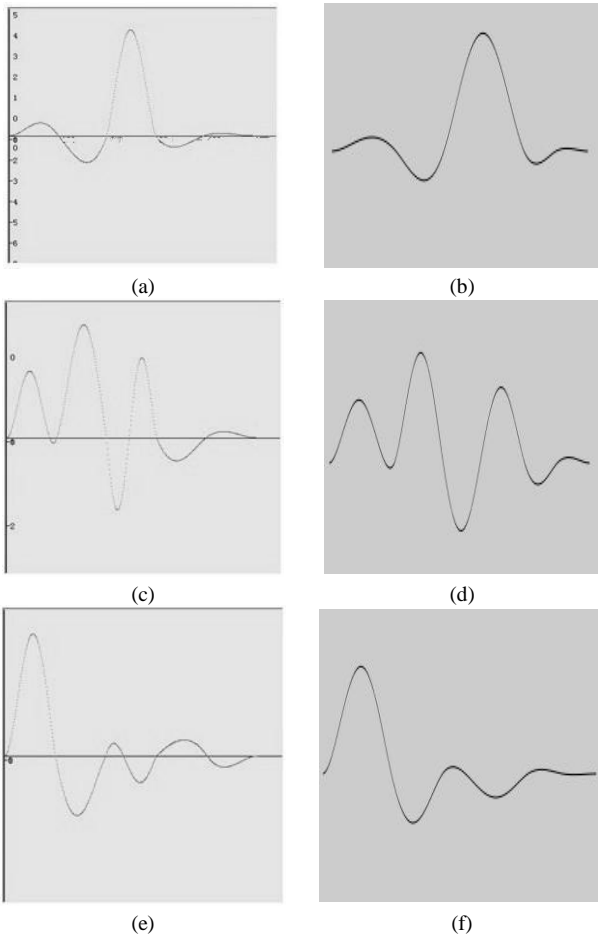


FIGURE 6 Vibration modals of unequal-span oil pipeline with rigid supports: (a) First-order modal in current model; (b) First-order modal in ABAQUS; (c) Second-order modal in current model; (d) Second-order modal in ABAQUS; (e) Third-order modal in current model; (f) Third-order modal in ABAQUS.

The computer code is used to investigate the vibration characteristics of the pipeline. A comparison between the calculation result of the first three-order modal of the pipeline and that from ABAQUS (in the ABAQUS 1000 B21 elements are used) is carried out. The first fifteen orders frequencies are listed in Table 3 and the vibration modals are shown in Figure 7.

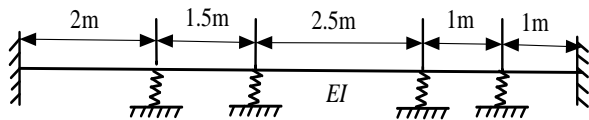


FIGURE 7 Unequal-span oil pipeline with elastic supports

It is shown in Table 3 that the first seven order frequencies respectively from the current method are very closed to those from ABAQUS and the maximum error between them is less than 4%. In Figure 6 the vibration modals from current method are also closed to those from ABAQUS. Compared the Table 2 with Table 3, it can be found that the natural frequencies of the pipeline with elastic supports are lower than those of the pipeline with rigid supports. This is consistent to the actual situation and easy to be understood because the stiffness of the

later is greater than the former. Thus, it can be seen the method in this paper is also efficient in the vibration modal analysis on unequal-span pipeline with elastic supports.

TABLE 3 The first five-order natural frequencies of the unequal-span pipeline with elastic supports

Frequency order	Current method	ABAQUS
1	14.821	15.289
2	41.371	40.764
3	80.798	78.866
4	133.388	128.70
5	199.214	189.61
6	270.408	260.78
7	300.408	315.45
6	270.261	260.78
7	300.841	315.45
8	370.408	341.26
9	457.750	430.26
10	594.268	526.89
11	643.151	630.78
12	726.448	739.83
13	870.557	845.57
14	954.677	945.87
15	959.739	973.71

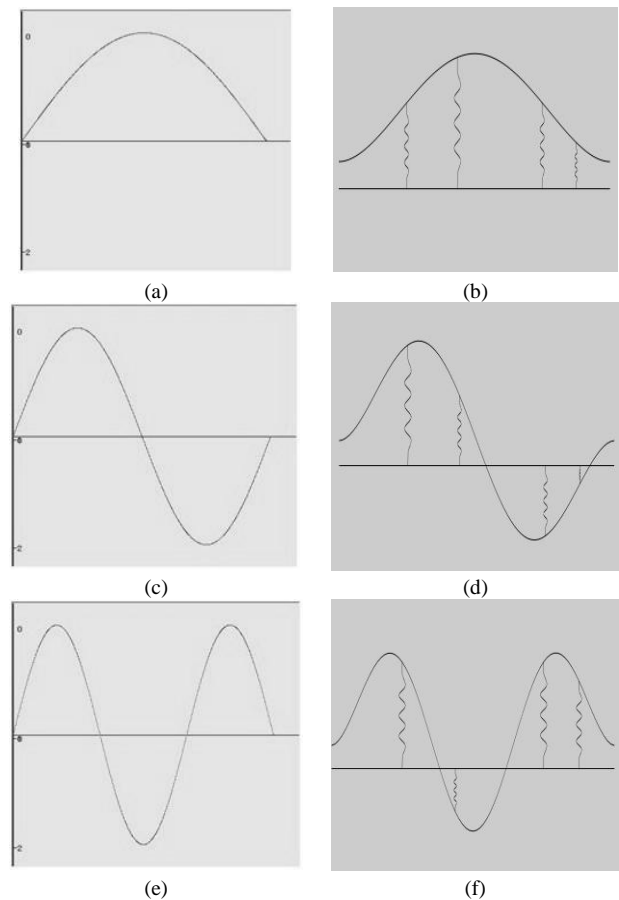


FIGURE 8 Vibration modals of unequal-span oil pipeline with elastic supports: (a) First-order modal in current model; (b) First-order modal in ABAQUS; (c) Second-order modal in current model; (d) Second-order modal in ABAQUS; (e) Third-order modal in current model; (f) Third-order modal in ABAQUS

5 Conclusions

An efficient vibration modal analysis model for unequal-span pipeline with different kinds of supports has been developed. The model is based on the earlier concepts of the matrix transfer, but has been extended to the pipeline with different spans and different kinds of supports. Two examples of unequal-span oil pipeline with respectively elastic supports and rigid supports are used to test the effectiveness of the model. The principal conclusions are as follows.

- (i) Compared with the ABAQUS software, the model not only shows good precise in the analysis on the low-order dynamic property, but also appears capability of computing high-order vibration modal.
- (ii) Commercial software based on finite element method may be more precise but is hard to be grasped by the ordinary designers. From this point, the model in this paper is more convenient for

References

- [1] Kaneko S, Kobayashi R, Watanabe T, Nakamura T 2012 *Journal of Physics: Conference Series* **382** 012048 6 pages
- [2] Yi Jia, Madeira R E, Madeira, Just-Agosto F 2005 Finite Element Formulation and Vibration Frequency Analysis of a Fluid Filled Pipe ASME 2005 International Mechanical Engineering Congress and Exposition, Recent Advances in Solids and Structures, Orlando, Florida, USA, November 5 – 11, 2005
- [3] Sollund H, Vedeld K 2012 A semi-analytical model for free vibrations of free spanning offshore pipelines, *Research Report in Mechanics*, No. 2 Dept. of Math., University of Oslo
- [4] Wang Guangyuan 1978 *Building vibration* Beijing: Science Press 168-75 (in Chinese)
- [5] Zhang Xide 1984 Plane-bending free vibration of continuous beam *Journal of Shandong Institute of Chemical Technology* 19-25 (in Chinese)

Liu Jun, He Xing, Liu Qinqyou, Naibin Jiang, Chen Huang practical application than commercial software, since there are not many parameters needed to be input in the model.

- (iii) Vibration characteristics analysis on multi-span line is a very important step in many engineering problems such as “Fluidelastic instability”, “Vibration induced by turbulence”. The realization of the model in this paper provides a choice concerned the vibration characteristics analysis model for these problems.

Acknowledgments

This research has been supported by the National Natural Science Foundation of China (Nos. 51105319 and 51274171), National Significant Science and Technology Special Sub-project of China (2011ZX05026-001-07) and National Key Laboratory of Nuclear Reactor System Design Technology Foundation of China.

- [6] Lian Huaying 1985 Research on the natural frequency of heat exchanger tube *Nuclear Power engineering* **6**(1) 28-33 (in Chinese)
- [7] Shen Huoming, Xiao Xinbin 2005 Vibration response of multi-span uniform beam under moving loads by using fitting beam mode function *Journal of Vibration and Shock* **24**(2) 27-32 (in Chinese)
- [8] Liu Haiping, Chang S C 2005 *Journal of Sound and Vibration* **281**(1-2) 155-69
- [9] Lin Hsien-Yuan, Tsai Yang-Chien 2007 Free vibration analysis of a uniform multispan beam carrying multiple spring mass systems *Journal of Sound and Vibration* **302** 442-56
- [10] Ye Mao, Tan Ping 2010 *Engineering Mechanics* **27**(9) 80-5
- [11] Hilal M Abu, Zibdeh H S 2000 *Journal of Sound and Vibration* **229**(2) 377-88
- [12] Li W L 2000 *Journal of Sound and Vibration* **237**(4) 709-25
- [13] Zheng Xiliang 2007 *Dynamic response analysis of the reverberation ray matrix method based on Timoshenko beam* PhD Thesis Zhejiang: Zhejiang University

Authors	
	<p>Liu Jun, born on May 8, 1980, Sichuan Province, PRC</p> <p>Current position, grades: PhD, associate professor</p> <p>Scientific interest: Engineering mechanics, structural dynamics, computational method and its application in engineering</p>
	<p>He Xing, born on March 15, 1990, Sichuan Province, PRC</p> <p>Current position, grades: Graduate student</p> <p>University studies: School of Civil Engineering and Architecture, Southwest Petroleum University, Chengdu, 610500, China</p> <p>Scientific interest: Structural engineering, Structural dynamics</p>
	<p>Qingyou Liu, born on October 12, 1965, Sichuan Province, PRC</p> <p>Current position, grades: PhD, professor</p> <p>Research interests: Engineering Mechanics Petroleum machinery, Pipe string mechanics, Rock breaking with bit and computer simulation</p>
	<p>Jiang Naibin, born on March 5, 1980, Sichuan Province, PRC</p> <p>Current position, grades: PhD, senior engineer</p> <p>University studies: graduated at Nuclear Power Institute of China, Ph.D degree in June 2013.</p> <p>Scientific interest: Structural mechanics in reactor technology</p>
	<p>Huang Chen, born on April 2, 1989, Sichuan Province, PRC</p> <p>Current position, grades: Graduate student</p> <p>University studies: School of Civil Engineering and Architecture, Southwest Petroleum University, Chengdu, 610500, China</p> <p>Scientific interest: Architectural and civil engineering, Structural dynamics</p>

The research of electromotor control based on optimized RBF neural network

Ni-qin Jing*, Lin-na Wang

Beijing polytechnic, Beijing 100015, China

Received 1 June 2014, www.tsi.lv

Abstract

RBF neural network suits to control electromotors, which have uncertainty and highly nonlinear systems. However, in practice, RBF neural network also have some obvious defects. For example, the strong dependence on the initial parameter and the poor quality of clustering algorithm. For the above defects, this paper is going to build an optimized RBF neural network through the combination of ant colony optimization algorithms, chaos ergodicity optimization theory and traditional K-means algorithm. On this basis, the optimized RBF neural network will be applied to PID control and then the dynamic performance of the electromotor will be simulationally tested by the designed PID controller. The simulation results show that in the control of electromotor, the optimized RBF neural network has the characteristic of high control accuracy and strong traceability and also it has the ability to guarantee electromotor control system with steady and dynamic performance.

Keywords: RBF neural network, ant colony optimization algorithms, chaos ergodicity optimization, chaos ant colony optimization algorithms, electromotor control

1 Introduction

The development of artificial neural network has experienced more than half a century. There emerges dozens of major structures of neural network. At present, the application of RBF neural network to PID control has become a hot issue, because RBF neural network has the ability to approximate any nonlinear function. It is able to learn parallel distributed processing and has a strong fault tolerance and robustness. Many training algorithms of RBF neural network support both online and offline training determine the network structure and the centre of the hidden layer units and width dynamically. And its learning speed is so fast that it is easy to model and control some complex nonlinear control systems.

However, RBF neural network is not omnipotent. It has a strong dependence on the set of the initial parameter. Once the initial parameter is given wrong, it will not get the optimized neural network structure. Another difficulty of RBF neural network is that the clustering quality of the traditional clustering algorithm is not high. Although the traditional clustering algorithm, like K-means algorithm, has a fast global searching speed, it is just a rough searching process, and in order to achieve optimal global searching effect, it needs to explore a new algorithm and then combine with it. Therefore, we introduce the ant colony algorithm and chaos colony algorithm based on the theory of chaos ergodicity to solve the initial parameter set and improve the traditional clustering algorithm.

We design the optimized RBF neural network PID controller under the Matlab in order to make the research of algorithm significant. We use this PID controller to make real-time status control of brushless direct current electromotor. And we test the feasibility and reliability of algorithm through the simulational experiment under the Matlab, making algorithm a new sublimation from theory to practice.

2 The Optimization of the RBF Neural Network

2.1 OPTIMIZED RBF NETWORK BASED ON ANT COLONY ALGORITHM

According to the advantages and disadvantages of K-means algorithm and ant clustering algorithm, we integrate these two algorithms to optimize RBF neural network jointly. By using the fast characteristic of K-means algorithm and the strong local searching ability of ant clustering algorithm, we design the improvement of RBF network clustering algorithm based on ant clustering algorithm. The main idea is using K-means algorithm and ant clustering algorithm to cluster the samples [5]. First, we use K-means algorithm to calculate the initial clustering centre of ant algorithm. Then we define the pheromone left by ant from sample X_i to the clustering centre $c_j(k)$ as τ_{ij} . The K in $c_j(k)$ is the K^{th} calculated clustering centre. The probability of ant (M) from X_i to $c_j(k)$ is:

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

$$P_{ij}^m = \frac{|\tau_{ij} \lambda_{ij}^\beta|}{\sum_{i \in \text{alloeedn}} \tau_{ij} \lambda_{ij}^\beta}, \quad (1)$$

where $i \in \text{alloeedn}$ is the samples which ant(M) can choose except the taboo table.

The followings are the updating equation of pheromone:

$$\tau_{ij}(k+1) = \rho \tau_{ij}(k) + \Delta \tau_{ij}, \quad (2)$$

$$\Delta \tau_{ij} = \sum_{k=1}^m \Delta \tau_{ij}^k, \quad (3)$$

$$\Delta \tau_{ij}^k = \frac{Q}{d_{ij}}, \quad (4)$$

where τ_{ij}^k is the pheromone of $X_i \rightarrow c_j(k)$, ρ is the persistence coefficient of strength, usually takes about 0.5-0.9 Q is the positive constant.

The followings are the steps of RBF neural network C and B confirmed by the detailed K-means algorithm and ant colony optimization algorithms [6].

Step1 Initialization algorithm: choose k different initial clustering centres

Step2 calculate the distance between all the sample inputs and clustering centre $\|X_i - c_j(k)\|$, $i=1, 2, \dots, n$; $j=1, 2, \dots, k$.

Step3 average the classified samples to get a new clustering centre $c_j(k+1)$. If $k+1 < N$, then do step2. If not, then do **Step4**.

Step4 initialization process: suppose when $k=0$, $NC=0$, $\tau_{ij}(0)=c$ (c is a constant), $\Delta \tau_{ij}=0$, $\lambda_{ij} = \frac{1}{d_{ij}}$ (an

expected factor) $\text{tabu}_m(s) = \phi$, $s=0$ (the initial stage of *taboo* table of every ant is empty), place m ants on n samples randomly. Then put the initial sample position of every ant in the current *taboo* table, and set $s=1$. S is *taboo* table index. Last, put the ant's initial city in the current *taboo* table.

Step5 repeat until the *taboo* table is full. Repeat $(n-1)$ times.

Step 6 calculate $d_{ij} = \|X_i - c_j(k)\|$, $i=1, 2, \dots, n$; $j=1, 2, \dots, k$.

Step7 according to the distance between every clustering centre above, then determine the base width vector of hidden nodes.

2.2 SIMULATIONAL EXPERIMENT

Based on the above improved clustering algorithm, we will approximate to a nonlinear system in order to compare the effect of the improved algorithm RBF network and the traditional RBF network. First of all, in

order to verify the feasibility of the algorithm [7], we can randomly choose 100 arrays between 0 and 1 as input samples. Overlap coefficient $\sigma=1$, number of clusters $k=8$, number of ants $m=30$, $Q=100$, iterations $NC=200$. Now through the simulational experiment of nonlinear system $f(x) = e^x + x \cdot x + \sin(x)$, we can get Figure1 and Figure2.

The experiment proves that it is feasible for ant algorithm to improve clustering algorithm and the effect of clustering is very good. After comparing these two figures, we can see that the ability of the improved algorithm RBF network to approximate nonlinear system is stronger, and also it is able to solve nonlinear system problems and put forward good solutions to the control of nonlinear system, such as electromotor.

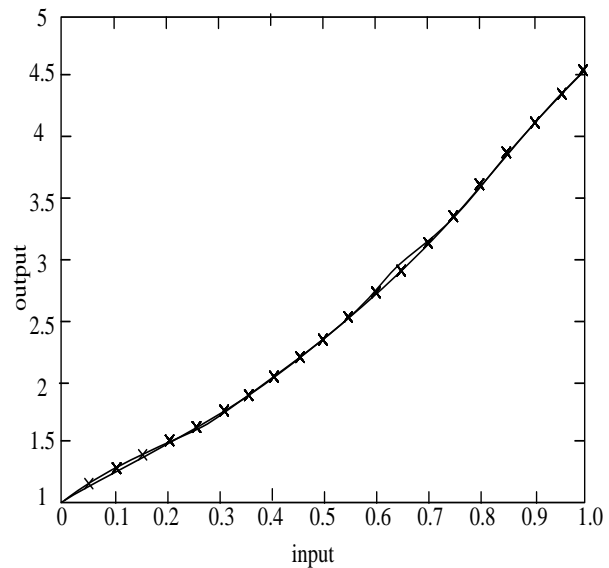


FIGURE 1 The functional effect figure of the improved algorithm RBF network approximation

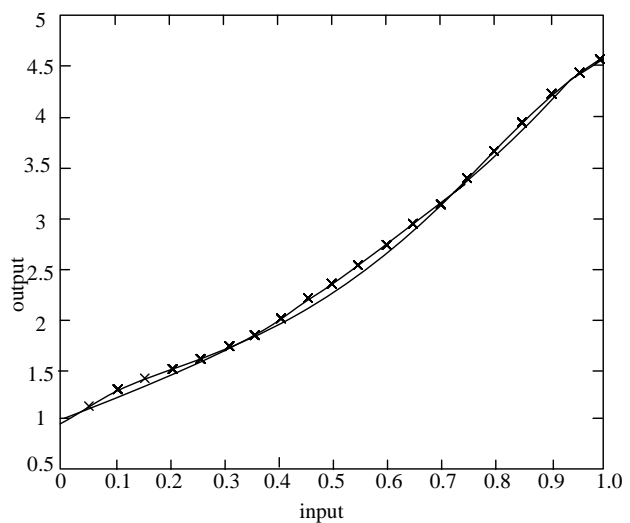


FIGURE 2 The functional effect figure of the traditional RBF network approximation

2.3 OPTIMIZED RBF NETWORK BASED ON CHAOS ANT COLONY ALGORITHM

In order to solve the problem of long time searching and local optimum of ant algorithm in optimization procedure, we adopt the good choice of the parameters in ant colony algorithm and the updated strategy of pheromone left in the path by the ant. We specifically introduce a chaotic disturbance in order to avoid the phenomenon of stagnation of ant colony algorithm in the optimization process. The chaos movement produced by Logistic mapping makes chaotic disturbance in pheromone of ant colony algorithm. The chaos movement produced by Logistic mapping can be defined like this:

$$h(n+1) = \mu h(n)[1-h(n)], \tag{5}$$

where $n=0,1,2,\dots$, $0 < h(n) < 1$, n means the iteration. When $\mu = 4$, it is in chaos state. $h(n)$ experiences all the states between 0 and 1.

The main idea of chaos ant colony algorithm is first using subtractive clustering method to cluster the samples [8]. Then use K-means algorithm to calculate the initial clustering centre of ant algorithm. Then define the pheromone left by ant from sample X_i to the clustering centre $c_j(k)$ as τ_{ij} . The probability of ant (M) from X_i to $c_j(k)$ is:

$$p_{ij}^m = \frac{|\tau_{ij} \lambda_{ij}^\beta|}{\sum_{i \in \text{alloeedn}} \tau_{ij} \lambda_{ij}^\beta} mQ, \tag{6}$$

where $i \in \text{alloeedn}$ is the samples which ant(M) can choose except the taboo table (as in (1)).

The followings are the updating equation of pheromone:

$$\left\{ \begin{array}{l} \tau_{ij}(k+1) = \rho \tau_{ij}(k) + \Delta \tau_{ij} \times 1.25 \times h(n) \\ \Delta \tau_{ij} = \sum_{k=1}^m \Delta \tau_{ij}^k \\ \Delta \tau_{ij}^k = \frac{Q}{d_{ij}} \end{array} \right. \tag{7}$$

$h(n)$ is the added chaotic disturbance of pheromone equation and n means iteration. This chaotic disturbance is in order to avoid the phenomenon of stagnation in searching optimization in ant colony algorithm and local extremum:

$$h(n+1) = 4h(n)[1-h(n)]. \tag{8}$$

In this equation, $h(n)$ meets the formula (5). τ_{ij}^k is the pheromone of $X_i \rightarrow c_j(k)$, ρ is the persistence coefficient of strength, usually takes about 0.5-0.9, Q is the positive constant. Detailed chaos ant colony algorithm to calculate RBF neural network node centre c and node base width b are the same as the above steps, so we will not talk about it here.

2.4 SIMULATIONAL EXPERIMENT

Through the optimization of RBF neural network optimized by chaos ergodicity and the optimization of RBF optimized by chaos ant colony algorithm, we are going to make an approximation to a nonlinear function in order to test the approximation performance made by RBF neural network, which is optimized by two algorithms. Thus, we can compare the advancement and feasibility of these two algorithms. First input the samples as the following methods: the number of input samples is 100, in which the input sample X_i must obey the uniform distribution of interval $[-4, 4]$. Then calculate the number of initial clustering centre by using subtractive clustering method. Suppose the number of ant(m) is 30, iterations $NC=200$, overlap coefficient $\sigma=1$. Through the approximation to the nonlinear function $f(x) = 1.1(1-x+2x^2)\exp(-x^2/2)$, we can get the simulation experiments Figure3 and Figure 4. From the comparison of these Figures, we can see that the training time needed for RBF neural network optimized by chaos ant colony algorithm is slightly more than RBF neural network optimized by chaos ergodicity, but its accuracy has a considerable improvement. We also can see that the nonlinear functional effects of RBF neural network optimized by chaos ant colony algorithm are better and more accurate.

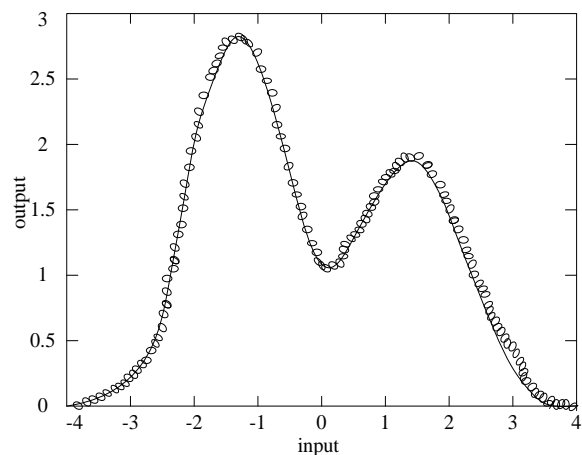


FIGURE 3 The nonlinear functional effect figure of RBF neural network optimized by chaos ant colony algorithm

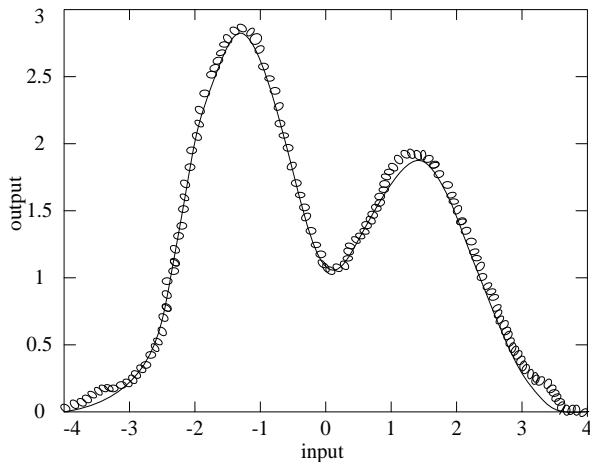


FIGURE 4 The nonlinear functional effect figure of RBF neural network optimized by chaos ergodicity

3 Brushless Direct Current Electromotor Control of the Optimized RBF Neural Network

3.1 PID CONTROL OF THE OPTIMIZED RBF NEURAL NETWORK

We adopt PID control of the optimized RBF neural network. The followings are the implementation of this control system.

We adopt the added PID control as the control system. The error control is:

$$e(k) = r(k) - y(k) . \tag{9}$$

PID input is:

$$\begin{cases} xc(1) = e(k) - e(k - 1) \\ xc(2) = e(k) \\ xc(3) = e(k) - 2e(k - 1) + e(k - 2) \end{cases} , \tag{10}$$

The control algorithm is:

$$u(k) = u(k - 1) + \Delta u(k) , \tag{11}$$

$$\begin{aligned} \Delta u(k) = & k_p (e(k) - e(k - 1)) + k_i e(k) \\ & + k_d [e(k) - 2e(k - 1) + e(k - 2)] . \end{aligned} \tag{12}$$

Take the performance index function $J = (yout(k) - y_m(k))^2 / 2$, in which $yout(k)$ is the k times output of the control system, $y_m(k)$ is the k times output of RBF neural network. When using RBF network to identify the system's Jacobi matrix, we can get:

$$\frac{\partial yout(k)}{\partial u(k)} \approx \frac{\partial y_m(k)}{\partial u(k)} = \sum_{j=1}^m \omega_j h_j \frac{c_{ji} - x_1}{b_j^2} . \tag{13}$$

In the above equation, x_1 is the control input $u(k)$.

3.2 SIMULATIONAL EXPERIMENT OF ELECTROMOTOR PERFORMANCE

In this paper, we use Matlab2007 as the test platform. After the comparison of neural network PID electromotor control improved by two algorithms and normal PID electromotor control, we can get the following result. In figure5, line1 represents the effect of normal PID control electromotor; line2 represents the effect of RBF neural network PID control electromotor optimized by chaos ant colony algorithm; line3 represents the effect of RBF neural network PID control electromotor optimized by ant colony algorithm.

From Figure 5, 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.

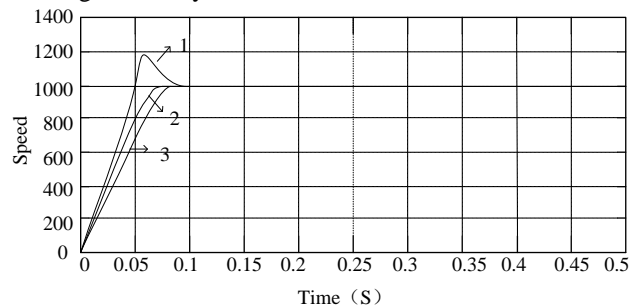


FIGURE 5 Operational process of electromotor under rated condition

From Figure 6, we can see that comparing to normal PID controller, RBF neural network PID controller improved by two algorithms have better speedy traceability when the given speed changes. As the rotated speed of electromotor changes, it can avoid overshoot shake, which is brought by normal PID control. Furthermore, the following speed are faster and that can reduce the energy consumption of the electromotor, so it is helpful to maintain the electromotor's stable working condition.

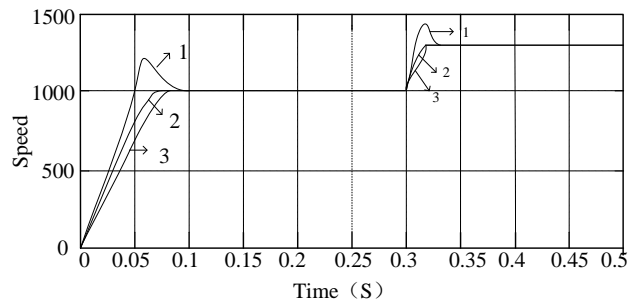


FIGURE 6 Electromotor's given speed from 1000 r/min to 1300 r/min

4 Conclusions

We can draw the following conclusions through the research of electromotor control based on optimized RBF neural network.

We use ant colony algorithm to improve k-means algorithm. Then we design the improvement of ant colony algorithm RBF network clustering algorithm through the combination of the theory of ant colony algorithm and the characteristic of k-means algorithm. This principle uses k-means algorithm when the initial clustering begins. Because this algorithm operates fast and suits global rough search [9]. Then we use ant colony to avoid local extremum in searching. These two combined clustering quality of clustering algorithm improve more obviously than the traditional clustering algorithm, and it is more suitable for the approximation of nonlinear system.

Through the characteristics of Logistic mapping, we design the specific methods to calculate the hidden layer neuron of RBF neural network. The optimized chaotic searching is very fast. It can avoid minimums and easily implement [10]. The function of chaos ant algorithm in RBF neural network is very successful, which better develops the ability of strong local optimization of ant colony algorithm.

We do the simulational experiments by the designed neural network PID controller. In the experiment, we simulate the appearance of load and disturbance. We control the working state of electromotor through RBF neural network, which is improved by RBF control in neural network PID controller by using two optimized algorithms. RBF neural network improved by two algorithms achieves good control effect on electromotors, which is proved by the experiments.

References

- [1] Maghami P G, Sparks D W Jr 2010 *IEEE Trans on Neural Networks* **11**(1) 113-23
- [2] Li T Y, Yorke J A 1975 *Amer Math Monthly* **82** 985-92
- [3] Robert J, Schilling J, Carroll J 2001 *IEEE Transactions on Neural Networks (SI045-9227)* **12**(1) 21-8
- [4] Chun Lianglin 2003 GA-based Multiobjective PID Control for a Linear Brushlees DC motor *IEEE Transactions on Mechstronics* **8**(1) 56-65
- [5] De Castro L N, Von Zuben F J 2002 *IEEE Transactions 011 Evolutionary Computation(SI089-778X)* **6**(3) 239- 51
- [6] Yang Y, Kamei M 2013 Clustering ensemble using swarm intelligence *IEEE Swarm. Intelligence Symposium Piscataway NJ: IEEE Service Centre* 35-7
- [7] Zheng Feng, Wang Qing Guo, Lee Tong Heng 2012 On the design of multivariable PID controllers via LMI approach *Automatic* **38** 517-26
- [8] Ismail Ossama, Bedwani Wassim 2001 Compliant motion control using variable structure PID control system *IEEE International Symposium on Intelligent Control* **5-7** 397-401
- [9] Rubaai Ahmed, Kotaru Raj 1999 *IEEE Trans. On System, Man, and Cybernetics* **29**(3) 460-74
- [10] Wu M 1989 *A lgorithms for Spanning Trees with Many Leaves and for Edge coloring of multigraph* Ph D Dissertation University of south Carolina

Authors



Niqin Jing, born in 1979, Shanxi Province of China

Current position, grades: lecturer

University studies: Master's degree was earned in major of software engineering, Beijing University of Technology in 2011.

Scientific interest: digital signal processing, Neural Network Algorithm, embedded system



Linna Wang, born in 1978, Henan Province of China

Current position, grades: lecturer

University studies: Master's degree was earned in major of signal and information processing, Tianjin University in 2011.

Scientific interest: digital signal processing

A universal tuning method of large dead-time system

Chang-liang Liu¹, Zeng-hui Ma^{1*}, Ping-an Kai²

¹State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (NorthChina Electric Power University), Beijing, 100052, China

²Energy Research Institute, National Development and Reform Commission, Beijing, 100010, China

Received 1 March 2014, www.tsi.lv

Abstract

The current approach of the PID controller design and tuning for the large dead-time system are almost based on the first-order time delay model, such as all PID algorithms and Ziegler function in MATLAB. And these algorithms generally only apply to the following system $0.1 \leq \tau/T \leq 2$. Therefore, the application effect of these algorithms in large dead-time system are not ideal. Based on the second-order system, this paper proposed a universal tuning method of PID controller for large dead-time process. By the introduction of the controller pre-coefficient K_f , this method makes the large dead-time system PID controller design and tuning simplistic. The fitting formula of controller pre-coefficient K_f was given in this paper. The method is simple, versatile, suitable for the object of $(\tau/T) \rightarrow +\infty$ and second-order, higher-order system, overcomes the limitations of traditional PID control algorithm in large dead-time system applications. The simulation results show that the method is correct, effective and has practical value.

Keywords: dead-time system, PID controller, parameter tuning, second-order system, pre-coefficient

1 Introduction

PID controller design and tuning for large dead-time system are recognized problem in control community. With the pure delay time τ increases, the control of dead-time system will become increasingly difficult [1, 9]. Due to the existence of pure time delay, the control system tends to instability, it is difficult to obtain good quality.

Now, using the powerful MATLAB language for PID controller design and tuning is very advanced and effective method. However, all PID algorithms in MATLAB are based on the first order time delay model. That is:

$$G(s) = \frac{k}{Ts + 1} e^{-\tau s} \quad (1)$$

If the first-order time delay model is more precise, the control effect of PID will be close to the control of the first-order time delay model. These algorithms are generally only applied to the following system $0.1 \leq \tau/T \leq 2$. These algorithms in MATLAB are not suitable for the controller design for large time delay system. Therefore, these algorithms have certain limitations on the scope of application [1, 2]. The well-known Ziegler-Nichols tuning formula mentioned in literature [3-4] is based on the first-order inertia and delay object. The tuning effect for large delay process using the

famous Ziegler-Nichols tuning formula is far from ideal. Similarly, the design of the Ziegler function in MATLAB is also based on the first order time delay model [1, 2]. In literature [5], the PI controller tuning for large dead-time process is also based on the first order time delay model. Nearly a hundred tuning methods and rules of PID controller for large dead-time process are showed in literature [6-7], but these tuning methods and rules are all based on the first order time delay model.

Therefore, whether the MATLAB or the Ziegler-Nichols tuning formula, now, all the PID controller-tuning methods of large dead-time process are based on the first-order time delay model, have certain limitations on the scope of application, and are powerless for large delay system.

The second-order system is the most representative system in the control. Generally, the controller for the second-order system is versatile [8]. Based on the second-order system, a simple and effective PID parameter tuning methods of large dead-time process was proposed. The method is simple, versatile and suitable for the object of $(\tau/T) \rightarrow +\infty$. This method solved the large delay control system design and tuning problem, which has been in existence. The simulation results show that the method is correct, effective and has practical value.

2 Control system design and tuning

There is a stable time-invariant second-order dead-time system:

* Corresponding author e-mail: mgh_1220@126.com

$$G(s) = ke^{-\tau s} / (as^2 + bs + 1) = e^{-\tau s} G_1(s), \quad (2)$$

where τ is the lag time, $e^{-\tau s}$ is the time lags, and

$$e^{-\tau s} = 1 - \tau s + (\tau s)^2 / 2! - (\tau s)^3 / 3! + \dots \quad (3)$$

In fact, $e^{-\tau s}$ is a non-minimum phase factor, equivalent to the controlled object has an infinite number of unstable zeros:

$$G_1(s) = k / (as^2 + bs + 1). \quad (4)$$

The designed control system is shown in Figure 1.

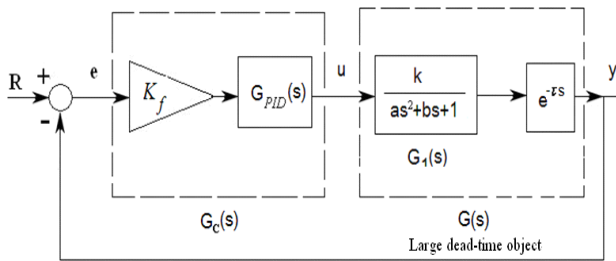


FIGURE 1 The PID control system of large dead-time process

Let

$$G_c(s) = G_{PID}(s)K_f. \quad (5)$$

The control system design methods can be divided into two steps.

Step 1: Design of PID controller

It is difficult to design the PID controller according to the equation 2 (second-order dead-time system). But it is easy to design the PID controller according to the equation 4 (stable time-invariant second-order system). According to the equation 4, by optimizing, the designed controller is:

$$G_{PID}(s) = [K_p + K_i / s + K_d s / (T_d s + 1)]. \quad (6)$$

There are many kinds of optimization methods can design the PID controller now. In order to avoid pure differential operation, we are often using the first-order lags to approximate pure differential link [1].

The PID controller parameters designed according to the equation 4(the object which not contain a pure time delay) is very easy tuning. We are trying to achieve the purpose of tuning the PID controller parameters by adjusting the controller pre-coefficient K_f when $e^{-\tau s}$ is added to the control system.

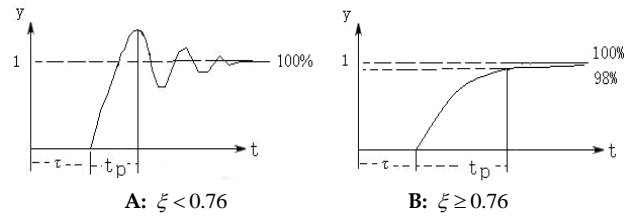


FIGURE 2 The step response of second-order dead-time system

Step 2: set and adjust the controller pre-coefficient

In Figure 2, A and B is the step response of the system shown in equation 2 in the case of damping $\xi < 0.76$ or $\xi \geq 0.76$. In Figure 2 A, t_p is the time of the step response overshoot to the first peak. In Figure 2 B, t_p is the time of the step response to achieve 98% steady-state value,

$$K_f = f(\tau / t_p) = f(h), \quad h = \tau / t_p. \quad K_f \text{ is a}$$

monotonically decreasing function about h . $0 < K_f \leq 1$, (if $\tau = 0$ then $K_f = 1$). Engineering practice shows that K_f must be reduced in order to ensure the stability and quality of the control system when the coefficient $h = \tau / t_p$ increases.

The closed-loop system output step response designed based on the equation 4(the object which not contain a pure time delay) is shown in figure 3A. Therefore, you can always find a value of K_f in the same PID parameters to make the following result holds. That is, the closed-loop system output step response of equation 2 (the object which contain a pure time delay) is the result of the closed-loop system output step response of equation 4 (the object which not contain a pure time delay) to the right pan τ . The closed-loop system output step response of the equation 2 is shown in Figure 3B.

Under the same PID parameters, each h is accordingly able to find the K_f , making no time delay object closed-loop output pan right τ to get the time delay object closed-loop output, which containing $e^{-\tau s}$.

$K_f = f(\tau / t_p)$, K_f can be obtained by curve fitting method.

$$K_f = f(h) = c_0 + c_1 h + c_2 h^2 + c_3 h^3 + c_4 h^4 + \dots, \quad (7)$$

or

$$K_f = f(h) = e^{-(c_0 + c_1 h + c_2 h^2 + c_3 h^3 + c_4 h^4 + \dots)}. \quad (8)$$

The nonlinear of computing K_f can compensate the impact of non-minimum phase factor $e^{-\tau s}$ in equation 2.

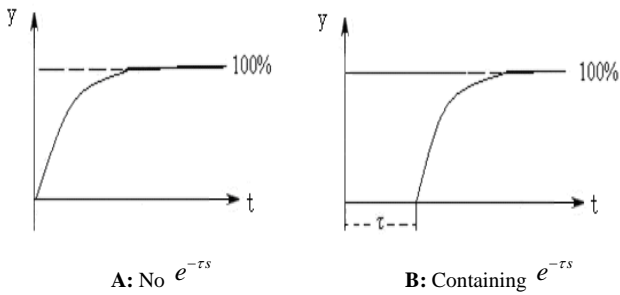


FIGURE 3 The step response of second-order closed loop system

3 Simulation study

The control system shown in Figure 1 was simulated as follows.

3.1 SIMULATION EXAMPLE 1

Input R is the step signal. The second-order time delay controlled object is shown in equation 2. Object parameters are $[k \ a \ b \ \tau]$. Simulation step is $t_s = 0.01s$.

- Let $[k \ a \ b \ \tau] = [1.6 \ 0.21 \ 0.37 \ 0]$; PID parameters is $[K_p \ K_i \ K_d \ T_d] = [0.45 \ 1.5 \ 0.27 \ 0.077]$; $K_f = 1$. The simulation results are shown in Figure 4. In Figure 4, the dotted line is the open-loop step response of controlled object; the solid line is the closed-loop output step response of controlled object (following the same).

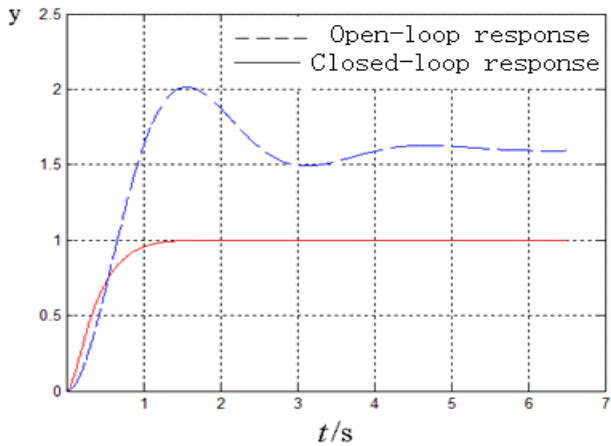


FIGURE 4 The simulation of example 1 ($\tau = 0$)

- Let $[k \ a \ b \ \tau] = [1.6 \ 0.21 \ 0.37 \ 0.4]$; In the same PID parameters, $K_f = 0.36$. The simulation results are shown in Figure 5. It can be seen, Figure 5 is the results of Figure 4 pan right 0.4s.

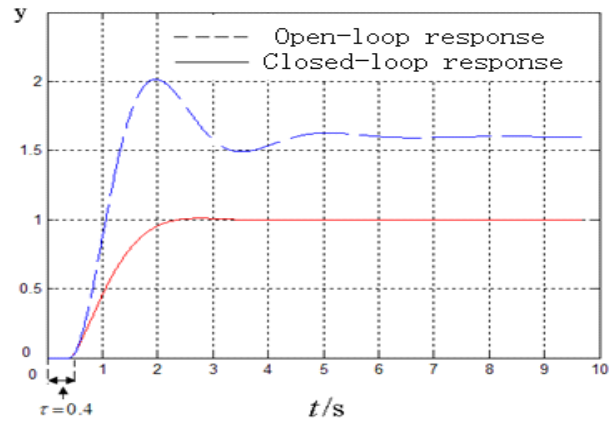


FIGURE 5 The simulation of example 1 ($\tau = 0.4$)

3.2 SIMULATION EXAMPLE 2

Input R is the step signal. The controlled object is $G(s) = ke^{-\tau s} / (Ts+1)$, the parameters is $[k \ T \ \tau]$, Simulation step is $t_s = 0.2s$.

- Let $[k \ T \ \tau] = [1.3 \ 100 \ 0]$, PI parameters are $[K_p \ K_i] = [1.37363 \ 0.0137363]$; $K_f = 1$; The simulation results are shown in Figure 6. In Figure 6, the dotted line is the open-loop step response of controlled object; the solid line is the closed-loop output step response of controlled object (following the same).

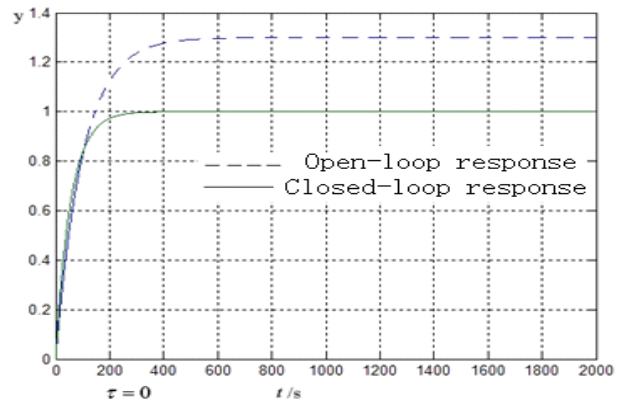


FIGURE 6 The simulation of example 2 ($\tau = 0$)

- Let $[k \ T \ \tau] = [1.3 \ 100 \ 100]$; In the same PI parameters, $K_f = 0.22$, the ratio of delay time τ and the time constant T is $\tau/T = 1$; The simulation results are shown in Figure 7. It can be seen, Figure 7 is the results of Figure 6 pan right 100s.
- Let $[k \ T \ \tau] = [1.3 \ 100 \ 400]$; In the same PI parameters, $K_f = 0.06$, $\tau/T = 4$; The simulation results are shown in Figure 8. It can be seen, Figure 8 is the results of Figure 6 pan right 400s.

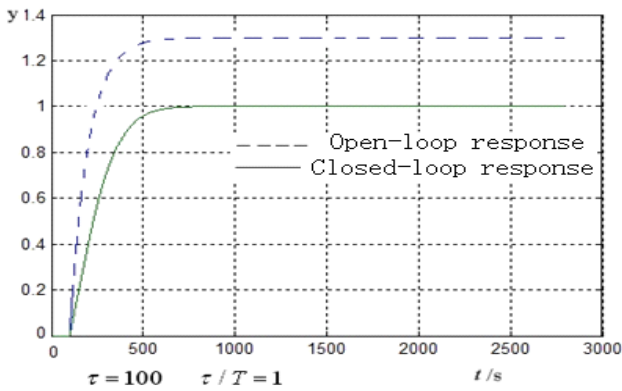


FIGURE 7 The simulation of example 2 ($\tau = 100$)

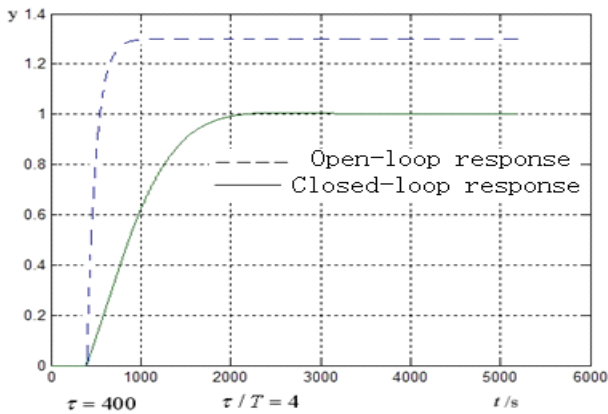


FIGURE 8 The simulation of example 2 ($\tau = 400$)

4) Let $[k \ T \ \tau] = [1.3 \ 100 \ 600]$; In the same PI parameters, $K_f = 0.04$, $\tau/T = 6$; The simulation results are shown in Figure 9. It can be seen, Figure 9 is the results of Figure 6 pan right 600s.

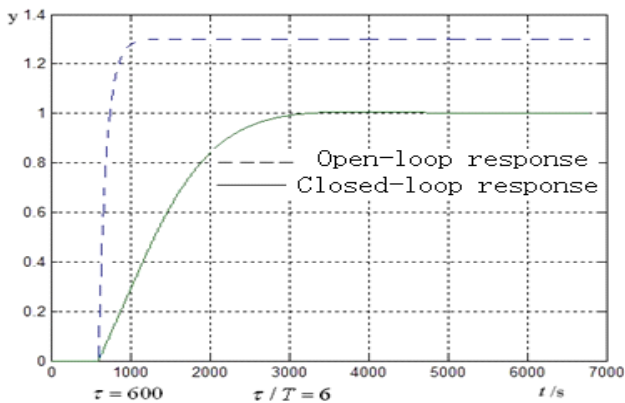


FIGURE 9 The simulation of example 2 ($\tau = 600$)

3.3 SIMULATION EXAMPLE 3

The higher-order object is $\frac{1.202}{(1+27.1s)^7}$ the higher-order object was reduced to the first-order delay system $\frac{1.202}{71.7s+1} e^{-118s}$.

Let $\tau=0$, $[K_p \ K_i] = [1.399 \ 0.03]$; $K_f = 1$; In the same PI parameters, let $\tau=118$, $K_f = 0.16$; The simulation results are shown in Figures 10 and 11. It can be seen, Figure 11 is the results of Figure 10 pan right 118s.

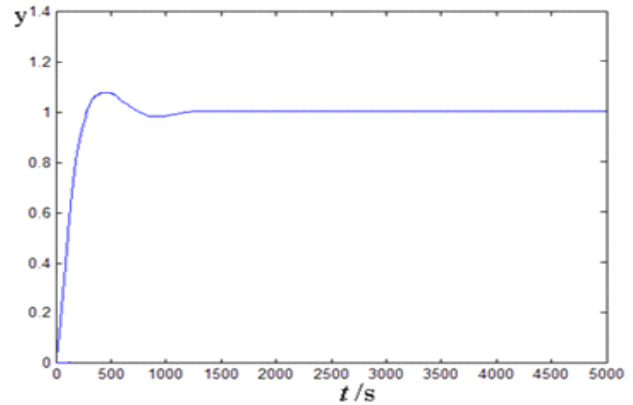


FIGURE 10 The simulation of example 3 ($\tau = 0$ 75% load)

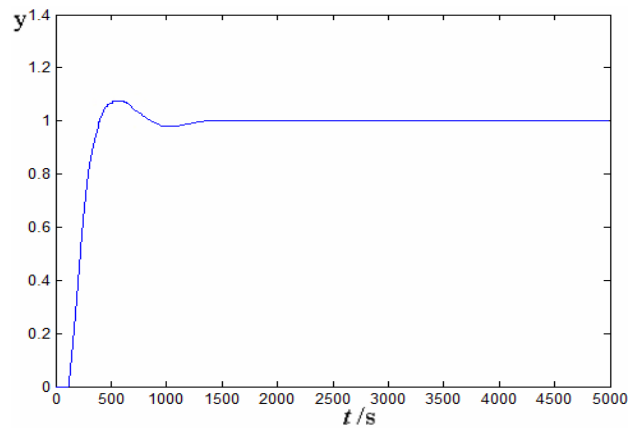


FIGURE 11 The simulation of example 3 ($\tau = 118$ 75% load)

It can be seen from the above simulation results that PID controller parameters can remain unchanged when large delay links $e^{-\tau s}$ is added to the control system, only need to adjust the system pre-coefficient can overcome the impact of dead-time and achieved the desired control effect.

The applications of the tuning method in second-order delay system and first-order delay system are shown as Example 1 and Example 2. It can be seen from example 2 that the method proposed in this paper is suitable for large time delay object, i.e. the object of $(\tau/T) \rightarrow +\infty$, and effectively overcomes the limitations of traditional design and tuning method in the application of large dead-time system. The applications of the tuning method in higher-order system, which can be reduced to first-order system are shown as Example 3.

5 Conclusions

The tuning method about large delay system in this paper has the following characteristics:

- 1) Simple and less adjustable parameters. When the controller design based on without delay stable system is completed, only need to adjust a parameter K_f can overcome the impact of the time delay. Adjusting a parameter K_f , which is equivalent to changing the PID controller four parameters $[K_p, K_i, K_d, T_d]$. So, makes the PID controller tuning method simplicity, practical and effective.
- 2) Has a strong commonality. The tuning method is suitable for not only the first-order delay system but also the second-order delay system(overshoot and no overshoot)and the higher-order delay system which can be reduce to the first-order or second-order system.
- 3) Wide range of applications. This method is suitable for the object of $(\tau/T) \rightarrow +\infty, K_f \rightarrow 0$; This is an improvement over the traditional design method. In the traditional method, the value of τ/T is limited to a lesser extent. It may be more effective if the method used in open-loop stepper control.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (Grant NO. 61203107).

References

- [1] Xue Ding-yu 2006 *Computer-aided Design of Control Systems-MATLAB and Its Application* 3 211-35 Beijing: Tsinghua University Press (in Chinese)
- [2] Visioli A 2001 Optimal tuning of PID controllers for integral and unstable processes *Proceedings of IEE D* 148(2) 180-4
- [3] Astrom K J, Hagglund T 2004 *Journal of Process Control* 14(6) 635-50
- [4] Hagglund T, Astrom K J 2002 *Asian Journal of Control* 4(4) 364-80
- [5] Xu Jiang-hua, Sun Rong, Shao Hui-he 2004 *Control and Decision* 1(19) 99-101
- [6] O'Dwyer A 1999 PI and PID controller tuning rules for time delay processes: a summary *Proceedings of the Irish Signals and Systems Conference Part1&2* 89-106
- [7] O'Dwyer A 2003 *Handbook of PI and PID controller tuning rules* London: Imperial College Press 123-45
- [8] Ho M T, Silva G J, Datta A, Bhattacharyya S P 2004 Real and Complex Stabilization: Stability and Performance *Proceedings of the American Control Conference* 4126-38
- [9] Soohee Han, Wook Hyun Kwon 2004 Receding Horizon Finite Memory Controls for Output Feedback Controls of Discrete-Time State Space Models *Proceedings of the 30th Annual Conference of the IEEE Industrial Electronics Society Busan: 2004* 1289-94
- [10] Kumar Padhy P, Majhi S 2009 Improved automatic tuning of PID controller for stable processes *ISA Transactions* 48(4) 423-7
- [11] Lee K N, Yeo Y K 2009 A new predictive PID controller for the processes with time delay *Korean Journal of Chemical Engineering* 26(5) 1194-1220
- [12] Normey-Rico J E, Camacho E F 2008 *Control Engineering Practice* 16(4) 407-28
- [13] Varol H A, Bingul Z 2004 A new PID tuning technique using ant algorithm *Proceedings of American Control Conference. Boston, Massachusetts, USA: IEEE, 2004* 2154-9
- [14] Coutinho D F, Fu M Y, de Souza C E 2010 *IEEE Transactions on Automatic Control* 55(3) 761-6
- [15] Malwatkar G M, Sonawane S H, Waghmare L M 2009 *ISA Transactions* 48(3) 347-53
- [16] Li Dazi, Zeng Fanyou, Jin Qibing et al 2009 Applications of an IMC based PID controller tuning strategy in atmospheric and vacuum distillation units *Nonlinear Analysis: Real World Applications* 10(5) 2729-39
- [17] Wang G L, Yab W W, Shao H H 2009 Multi-objective optimization based on Genetic Algorithm for PID controller tuning *Journal of Harbin Institute of Technology* 16(1) 71-4
- [18] Lai C L, Hsu P L 2010 *IEEE Transactions on Industrial Informatics* 6(1) 73-80
- [19] Zhang Yan, He Yong, Wu Min 2009 Delay-dependent robust stability for uncertain stochastic systems with interval time-varying delay *Acta Automatica Sinica* 35(5) 577-82
- [20] Normey-Rico J E, Camacho E F 2009 Unified approach for robust dead-time compensator design *Journal of Process Control* 19(1) 38-47

Authors	
	<p>Chang-liang Liu</p> <p>Current position, grades: Professor, PhD North China Electric Power University(China) Research interests: control simulation, system modelling, advanced control theory</p>
	<p>Zeng-hui Ma</p> <p>University studies: PhD student of North China Electric Power University(China)</p>
	<p>Ping-an Kai</p> <p>Current position, grades: Professor, Energy Research Institute, National Development and Reform Commission(China) Research interests: advanced control theory and applications, modelling and optimization</p>

Using grey-weighted Markov chain model to predict the quantum of highway passenger transport

Wenjing Li*

School of Science, Shandong Jiaotong University, No.5 Jiaoxiao Road, Jinan, Shandong, China, 250023

Received 1 May 2014, www.tsi.lv

Abstract

The grey-weighted Markov model is a prediction model integrating the advantages of grey model and Markov chain model. It can be applied to predict the highway passenger transport quantum. Compared with grey model, the grey-weighted Markov chain model improved the precise of prediction, so the combined model was more appropriate for the prediction of highway passenger transport. Based on the original data of highway passenger transport quantum from 2001 to 2011, the passenger transport quantum in 2012 was predicted with grey-weighted Markov chain model.

Keywords: Grey model, weighted Markov chain, passenger transport quantum, prediction

1 Introduction

Grey system theory was brought forward and developed by Deng Julong in 1982 [1]. Grey model, i.e. GM, is a useful model to solve some problems with short time series, less statistical data, or incomplete information. But GM has a worse fitting and precision for the long term prediction or data series with great random wave (Deng 1989) [2]. Markov chain is a restricted class of stochastic process with finite or denumerable state-space. Their main property is that the joint distributions of the involved variables are fixed by the transition matrices of the process and the distribution of the initial variable. The transition matrices show the internal wave rule of state-space by which we can fix random errors of GM, therefore we can combine GM and Markov chain to predict value. Actually, the combined prediction models have widely appeared both in pure and applied mathematics, and have many applications in science and technology [3-8].

The highway passenger transport quantum is the number of passengers conveyed through the highway transportation system in some a period. The passenger transport quantum is an important data index reflecting the level of the transportation serves for national economy and people's living. It is also the index of programming the passenger transport quantum and studying the scale and the pace of the development of transportation. There are several methods to predict the passenger transport quantum such as the experts' experiences prediction method, the algorithm average method, the liner regression method, and BP neural network model etc, every method has advantages and disadvantages (Wang 2007) [9-10]. This paper mainly apply grey-weighted Markov chain model to predict the

highway passenger transport quantum. Compared with GM, the combined model improved the prediction precise.

2 Establishment of the model

2.1 GM (1, 1)

Step 1: Accumulate original series

$x^{(0)}(k)$ is the highway passenger transport quantum of the k th year, then original sequence is set:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(N)\}. \quad (1)$$

New sequence is created and defined as follows:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(N)\} \\ = \left\{ x^{(0)}(1), \sum_{i=1}^2 x^{(0)}(i), \dots, \sum_{i=1}^N x^{(0)}(i) \right\}. \quad (2)$$

Step 2: Establish the equation

GM (1,1) is a prediction model with an order and one variable. Its whitened equation is as follows:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \quad (3)$$

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

Where “ a ” is named evolution grey coefficient, “ u ” is named end genesis-control grey value, $U = \begin{bmatrix} a \\ u \end{bmatrix}$. \hat{U} is estimated by the method of GLS:

$$\hat{U} = (B^T B)^{-1} B^T Y, \tag{4}$$

where $B = \begin{pmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(N-1) + x^{(1)}(N)) & 1 \end{pmatrix}$,

$$Y = [x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(N)]^T.$$

We can obtain the estimation of “ a ” and “ u ” by \hat{U} , and then obtain the time-response Eq. which is as follows (Deng 1990):

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}, \quad k = 1, 2, \dots, \tag{5}$$

when $k = 1, 2, \dots, N-1$, $\hat{x}^{(1)}(k+1)$ is simulated value, when $k \geq N$, $\hat{x}^{(1)}(k+1)$ is predicted value.

Step 3: $x^{(0)}$ is deoxidized by $\hat{x}^{(1)}$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, 3, \dots, \tag{6}$$

where $\hat{x}^{(0)}(1) = x^{(0)}(1)$, then $\hat{x}^{(0)}(k+1)$ is obtained:

$$\hat{x}^{(0)}(k+1) = (1 - e^a)x^{(0)}(1) - \frac{\hat{u}}{\hat{a}}e^{-\hat{a}k}. \tag{7}$$

Step 4: The test of precision of GM (1,1)
Error is defined as:

$$E(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), \quad k = 1, 2, \dots, N. \tag{8}$$

Relative error is defined as:

$$e(k) = \frac{E(k)}{x^{(0)}(k)} \times 100\%, \quad k = 1, 2, \dots, N. \tag{9}$$

Deviation of original series is defined as:

$$S_1^2 = \frac{\sum_{k=1}^N [x^{(0)}(k) - \bar{x}]^2}{N}, \tag{10}$$

where $\bar{x} = \frac{\sum_{k=1}^N x^{(0)}(k)}{N}$.

Deviation of error series is defined as:

$$S_2^2 = \frac{\sum_{k=2}^N [E(k) - \bar{E}]^2}{N-1}, \tag{11}$$

where $\bar{E} = \frac{\sum_{k=2}^N E(k)}{N-1}$.

Posterior ratio is calculated:

$$C = \frac{S_2}{S_1}. \tag{12}$$

The probability of error is calculated:

$$P = P\{|E(k) - \bar{E}| < 0.6745S_1\}. \tag{13}$$

The quality of prediction can be judged by table 1.

TABLE 1 The standard of the quality of GM (1, 1)

Grade	P	C
Good	$P > 0.95$	$C < 0.35$
Acceptable	$0.80 \leq P < 0.95$	$0.35 \leq C < 0.50$
Conceded acceptable	$0.70 \leq P < 0.80$	$0.50 \leq C < 0.65$
Unacceptable	$P \leq 0.70$	$C \geq 0.65$

2.2 GREY-WEIGHTED MARKOV CHAIN MODEL

Step 1 We establish the standard of the grade of relative error, i.e. set up the status bar of Markov chain according to the solution of GM(1,1).

Step 2 The grate of relative error sequence is ascertained by the standard.

Step 3 Autocorrelation coefficient r_k is calculated as follows:

$$r_k = \frac{\sum_{l=1}^{N-k} (x_l - \bar{x})(x_{l+k} - \bar{x})}{\sum_{l=1}^N (x_l - \bar{x})^2}. \tag{14}$$

r_k is the autocorrelation coefficient which step size is k , x_l is the l th relative error.

\bar{x} is the average of relative error, N is the length of highway passenger transport quantum sequence.

Step 4 Autocorrelation coefficient standardization

The weight of the Markov chain with step size k is defined as:

$$w_k = \frac{|r_k|}{\sum_{k=1}^m r_k}, \tag{15}$$

(m is the max order which the prediction model need).

Step 5 According to the transitive state-space of relative error series; we can obtain the transition probability matrix of different step sizes. Noting the original sequence and the corresponding transition probability matrix, we can predict the state possibility $p(k)$ of the relative error of the highway passenger transport quantum, where k is the step size, $k = 1, 2, \dots, m$.

We sum m weighted possibilities of the same state-space and regard the result as the possibility of the prediction, furthermore, the “ i ”, which corresponds $\max\{p(i), i \in E\}$ is the grade of relative error of prediction passenger transport quantum. We can fix the result of GM by the median of interval of relative error. Adding the final result to original sequence and repeat the

TABLE 2 The error and relative error of every year

Year	Actual value (ten thousand)	Simulated value (ten thousand)	Error (ten thousand)	Relative error (%)
2001	70497	70497.0000	0	0
2002	74626	56825.5172	17800.4828	23.8529
2003	75492	68203.3143	7288.6857	9.6549
2004	89388	81859.2124	7528.7876	8.4226
2005	98485	98249.3405	235.6595	0.2393
2006	109472	117921.1556	-8449.1556	-7.718
2007	123963	141531.7279	-17568.7279	-14.173
2008	168675	169869.6888	-1194.6888	-0.708
2009	234234	203881.5720	30352.428	12.9582
2010	248720	244703.4294	4016.5706	1.6149

According to table 2, the average of relative errors from 2001 to 2010 is 3.9033, the standard deviation is 11.60. The relative error is graded as follows:

The state-space of every year is ascertained according to the table 3:

above steps, we can predict the passenger transport quantum of next year.

3 Example analysis

The paper takes the highway passenger transport quantum of Shandong province from 2001 to 2010 for an example and uses grey-weighted Markov chain model to simulate or predict (the data source from Chinese Stat. Annual).

Step 1: With the program of GM(1,1), we get $a=-0.1825, u=38931.3719$, then get the simulation of the highway passenger transport quantum from 2001 to 2010 (table 3.1) and the prediction of the passenger transport quantum in 2011: $\hat{x}(2011) = 293698.7771$ (ten thousand).

The posterior ratio is calculated: $C = S_2/S_1 \approx 0.0225 < 0.35$, the probability of error is calculated: $P = P\{|E(k) - \bar{E}| < 0.6745S_1\} = 1 > 0.95$.

According to table 1, the quality of GM(1,1) is “Good”.

Step 2: We optimize the prediction by the mean of interval of relative error of weighted Markov chain.

Relative error is calculated as follows:

TABLE 3 The grade interval of relative error

Relative error state-space	The standard of grade	Relative error interval
1	$\bar{x} - 2.0s \leq x \leq \bar{x}$	$-19.29 \leq x \leq 3.90$
2	$\bar{x} < x \leq \bar{x} + 1.0s$	$3.90 < x \leq 15.50$
3	$\bar{x} + 1.0s < x \leq \bar{x} + 2.0s$	$15.50 < x \leq 27.10$

TABLE 4 The grade of relative error of every year

Year	Relative error	Grade	Year	Relative error	Grade
2001			2006	-7.718	1
2002	23.8529	3	2007	-14.173	1
2003	9.6549	2	2008	-0.708	1
2004	8.4226	2	2009	12.9582	2
2005	0.2393	1	2010	1.6149	1

Autocorrelation coefficients of different orders and weights of different step-sizes of Markov chain are calculated as follows:

TABLE 5 The autocorrelation coefficient and weight of different orders

State-space	1	2	3
r_k	1.4931	-0.0456	-1.0079
w_k	0.5863	0.0179	0.3958

From table 4, we can get the transition probability matrixes of different step sizes as follows:

$$p_1 = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 2/3 & 1/3 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad p_2 = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$p_3 = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

We can predict the state-space of relative error of highway passenger transport quantum in 2011 according to the data and the corresponding transition probability matrixes of 2008~2010. The results are as follows:

TABLE 6 The prediction of the range of relative error in 2011

Original year	State	Step size	Weight	1	2	3	Resource
2010	1	1	0.5863	0.75	0.25	0	p_1
2009	2	2	0.0179	1	0	0	p_2
2008	1	3	0.3958	2/3	1/3	0	p_3
Sum				0.4577	0.2772	0	

According to table 6, $\max\{p_i, i \in E\} = 0.4577$, the corresponding state space $i=1$, i.e. the relative error state-space is 1 in 2011. We fix the result of GM by the median of interval of relative error. Finally, we get the prediction of the highway passenger transport quantum of Shandong province in 2011: $\hat{x}(2011) = 272700$ (ten thousand).

The real highway passenger transport quantum of Shandong province in 2011 is 250469(ten thousand). It is obvious that the relative error of grey-weighted Markov combined model is smaller than the relative error of GM (1, 1), see table 7:

TABLE 7 Compare the precision of GM with Grey-weight Markov model

Year	Model	Actual value (ten thousand)	Simulated value (ten thousand)	Relative error (%)
2011	GM(1,1)	250469	293698	-17.26
	Grey-weight Markov	250469	272700	-8.88

In the same way, we can estimate other years, e.g. we compare the results of two models applied on the

prediction of passenger transport quantum in 2009 as follows:

TABLE 8 Compare the precision of GM with Grey-weight Markov model

Year	Model	Actual value (ten thousand)	Simulated value (ten thousand)	Relative error (%)
2009	GM(1,1)	234234	203882	12.96
	Grey-weight Markov	234234	225783	3.61

Therefore, we choose grey-weighted Markov model to predict the highway passenger transport quantum of Shandong province in 2012 as follows:

TABLE 9 The prediction of the highway passenger transport quantum in 2012

Year	Model	State	Predicted value (ten thousand)
2012	GM(1,1)		319387
	Grey-weight Markov	1	296553

$\hat{x}(2012) = 296553$ (ten thousand).

4 Conclusions

This paper compares the performances of GM (1, 1) and grey-weighted Markov chain method in passenger traffic quantum prediction. It is shown, that the latter model adjusts the error of GM, and then improves the prediction precision. However, the transition possibility matrix lies on the frequency, so the precision of grey-weighted Markov chain method is affected by the length of time series, we can obtain more accurate result with longer original time series.

Acknowledgments

The paper is supported by the Natural Science Fund of Shandong Jiaotong University (Z201338).

References

- [1] Deng J L 1982 The control problems of grey systems *Systems Control Letters* **5** 288-94
- [2] Deng J L 1989 Introduction to grey system theory *The Journal of Grey System* **1** 1-24
- [3] Hasantash, Gholam Hossein, Mostafaei, Hamidreza, Kordnoori, Shaghayegh 2012 Modeling the Errors of EIA's Oil Prices and Production Forecasts by the Grey Markov Model *International Journal of Economics & Financial Issues (IJEFI)* **2** 312-9
- [4] Erdal Kayacan, Baris Ulutas, Okyay Kaynak 2010 Grey system theory-based models in time series prediction *Expert Systems with Applications* **37** 1784-9
- [5] Shaghayegh Kordnoori, Hamidreza Mostafaei 2011 Grey Markov Model for Predicting the Crude Oil Production and Export in Iran *International Journal of Academic Research* **3** 1029-33
- [6] Chen Li-Hui, Guo Tsuei-Yang 2011 Forecasting financial crises for an enterprise by using the Grey Markov forecasting model *Quality & Quantity* **45** 911-22 (*In Chinese*)
- [7] Kumar Ujjwal, Jain V K 2010 Time series models (Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India *Energy* **35** 1709-16
- [8] Hsu Yen-Tseng, Liu Ming-Chung, Yeh Jerome, Hung Hui-Fen 2009 Forecasting the turning time of stock market based on Markov-Fourier grey model *Expert Systems with Applications* **36** 8597-603 (*In Chinese*)
- [9] Wang Zhuo, Jia Li-Min, Qin Yong, Wang Yan-hui 2007 Railway Passenger traffic volume prediction based on neural networks *Applied Artificial Intelligence* **21** 1-10 (*In Chinese*)
- [10] Deng Wu, Li Wen, Yang Xin-hua 2011 A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction *Expert Systems with Applications* **38** 4198-205 (*In Chinese*)

Authors**Wenjing Li**

Current positions, grade: an associate professor of Shandong Jiaotong University.

Scientific interest: Markov Chain, Grey model and other mathematics methods.

Publications: have appeared in some well-known conference proceedings and international journals. **Experience:** leading some research projects supported by the university and Chinese government, such as the Natural Science foundation, the project of Statistic, etc.

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

Received 1 May 2014, www.tsi.lv

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: Modelling in SolidWorks, interface processing, birth-death element, APDL, simulation analysis in 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 defense 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

*Corresponding author e-mail: 1694440@qq.com

widely used in civil engineering, traffic, water conservancy, railway, petroleum chemical industry, aerospace, machinery manufacturing, national defense, 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 pre-processing 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 sidewall modelling. Firstly, selecting sketch map and using sketch-rendering tools to draw L-shaped cross-section of base plate and sidewall. 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 sidewall 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 Figure2.

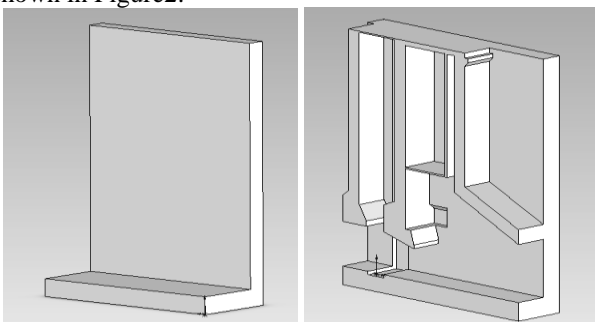


FIGURE 1 3-d L-shaped model FIGURE 2 Internal structure model

(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.

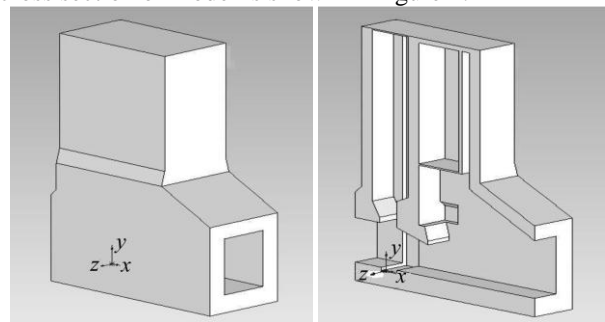


FIGURE 3 Integrated model

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 Defeating" or not are given in Figure 5 and Figure 6. Both of models can be pre-processed in ANSYS in practice.

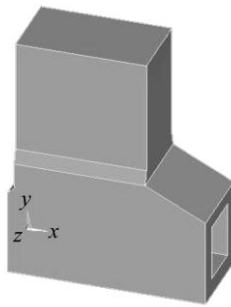


FIGURE 5 With defeaturing

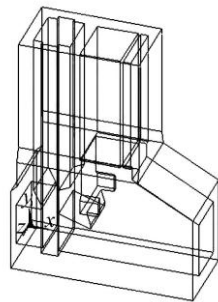


FIGURE 6 Without defeaturing

3 Analysis of thermal field of intake tower

There are 5 products of ANSYS to thermal analysis as shown in Figure 7, which can be classified into ANSYS/Multiphysics, ANSYS/Mechanical, ANSYS/Thermal, ANSYS/FLOTRAN and ANSYS/ED.

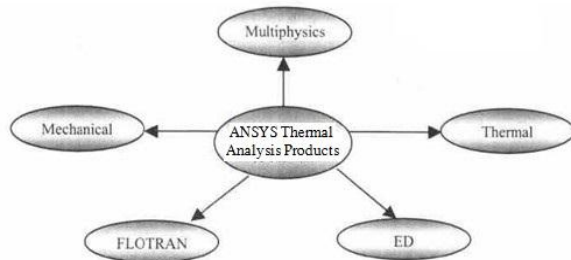


FIGURE 7 ANSYS thermal analysis products

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, autogenous 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 which will be killed
EKILL, ...! Killing the elements

ESEL,... !Selecting the elements which 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]:

$$\frac{\partial T}{\partial \tau} = \frac{\lambda}{c\rho} \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + \frac{\partial \theta}{\partial \tau} \quad , \quad (1)$$

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:

$$I^e(T) = \iiint_{\Delta R} \left\{ \frac{1}{2} \alpha \left[\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right] + \iiint_{\Delta R} \left(\frac{\partial T}{\partial \tau} - \frac{\partial \theta}{\partial \tau} \right) T \right\} dx dy dz + \iint_{\Delta D} \bar{\beta} \left(\frac{1}{2} T^2 - T_a T \right) ds \quad (2)$$

where ΔR is a subfield of unit e, the thermal diffusivity $\alpha = \frac{\lambda}{c\rho}$, $\bar{\beta} = \frac{\beta}{c\rho}$, 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]:

$$Q(t) = 71610 [1 - \exp(-0.36t)] \quad (3)$$

where t is the pouring time, and Q is the hydration heat. The relation between Q and θ is written as:

$$\frac{\partial \theta}{\partial \tau} = \frac{Q}{c\rho} \quad (4)$$

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 $45kJ/(m^2 \cdot h \cdot ^\circ C)$ and $10kJ/(m^2 \cdot h \cdot ^\circ C)$, 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, $T = 18^\circ C$;
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 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		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

which underwater is 24.413 °C, occurring on the 285th day of the total construction and impounding periods.

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				

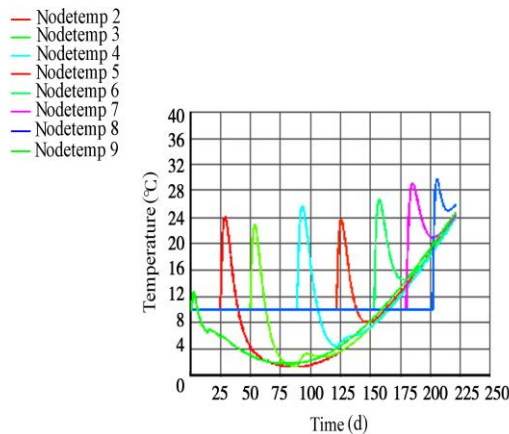


FIGURE 8 Maximum temperature curves

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.

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

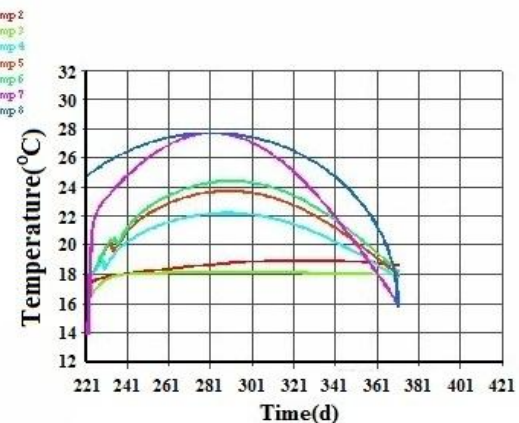


FIGURE 9 Maximum temperature curves

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/m ³)	Coefficient of linear expansion (1/°C)	Poisson ratio
Concrete	2 447.5	9×10 ⁻⁶	0.167

The elasticity modulus is written as:

$$E_t = 3.6 \times 10^{10} [1 - \exp(-0.40t^{0.34})] \tag{12}$$

The creep effect is considered in analysis of temperature stress, and the formula of the creep degree is written as:

$$C = \left[0.23(1 + 9.2t^{-0.45})(1 - e^{t_1}) + 0.52(1 + 1.17t^{-0.45})(1 - e^{t_2}) \right] \times 10^{-10} / 3.60 \quad (13)$$

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,

$$E = E_1 / (1 + CE_1) \quad (14)$$

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.

In Figure 10, 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,

$$\sigma_t = 0.232 \times 10^6 \left\{ 33.5 \left[1 + 0.2 \times \ln(t/28) \right] \right\}^{\frac{2}{3}} \quad (15)$$

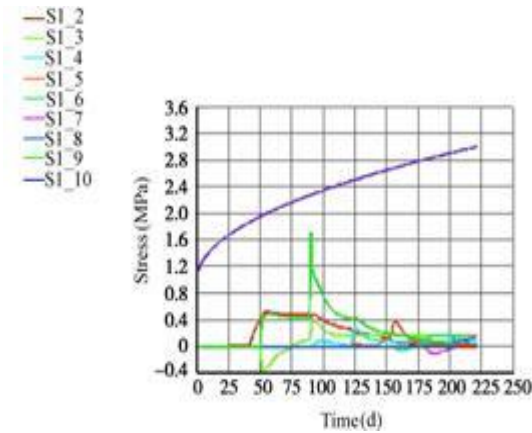


FIGURE 10 Maximum stress curves

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.

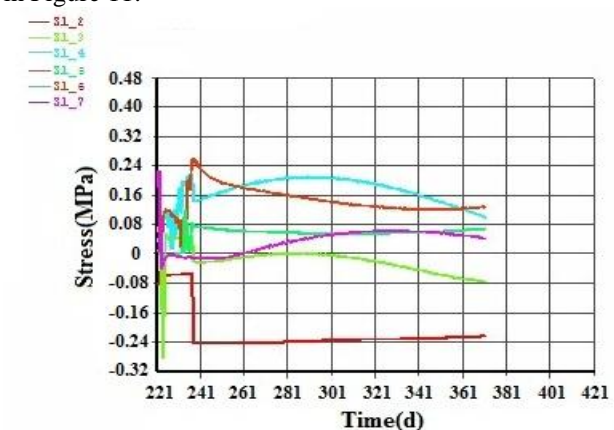


FIGURE 11 Maximum stress curves

In Figure 11, 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

The authors wish to thank the National Natural Science Foundation of China for contract 51279064 and 41061046, and Open Fund of China Institute of Water Resources and Hydropower Research of China for contract IWHR-SKL-201110, under which the present work was possible.

References

- [1] Ashida F, Tauchert T R 1998 *Journal of Applied Mechanics* **65** 367-73
- [2] Kawaguchi T, Nakane S 1996 *ACI Materials Journal* **93** 96-101
- [3] Lin J Y, Cheng T F 1997 *Numerical Heat Transfer* **32** 187-203
- [4] Wu Y, Luna R 2001 *Finite Elements in Analysis and Design* **37** 97-106
- [5] Wu J G, Li X J, He K F 2012 *Journal of Computers* **7** 2208-15
- [6] Li X J, Wang K, Jiang L L 2012 *Journal of Software* **7** 959-65
- [7] Zhang J 2005 *Chinese Quarterly of Mechanics (Shanghai)* **26** 257-62 (in Chinese)
- [8] Du P, Liu S X, Tan G Z, Liu K X 2012 *Journal of Liaoning Technical University(Natural Science) Fuxin* **31** 526-30 (in Chinese)
- [9] Zhao Y W, Geng D X, Liu X M 2012 *Journal of Computers* **7** 2742-9
- [10] Jin K K, Kook H K, Yang J K 2001 *Computers and Structures* **79** 163-71
- [11] Liu L J, Ren J P 2005 *Mechanical Management and Development, (Datong)* **20** 74-75 (in Chinese)
- [12] Prabhakar S, and Henderson M R 1992 *Computer-Aided Design* **24** 381-93
- [13] Rajan G, Venkat D 1998 *Computer-Aided Design* **30** 863-73
- [14] Gao Y 2010 *Journal of Yangzhou Polytechnic College* **14** 22-24 (in Chinese)
- [15] Li X, Huang Z F, Li H T 2013 *Journal of Software* **8** 2517-21
- [16] Gao P, Li P H, Bao K M 2012 *Journal of China Three Gorges university (Yichang)* **34** 29-33(in Chinese)
- [17] Chen G R, Xu W T, Yang Y., Li K 2012 *Chinese Journal of Computational Physics* **29** 411-6 (in Chinese)
- [18] Li X F 2012 *Concrete* **22** 127-9(in Chinese)
- [19] Chen S K, Zhang Y P, Guo L X, Wang H. B, Xie Z Q 2012 *Concrete* **22** 10-6 (in Chinese)
- [20] Liu X H, Ma G., Chang X L, Zhuo W 2012 *Engineering Mechanics* **29** 159-164
- [21] Ma C C, Li S. Y, Zhao L J, Liu C Y 2012 *Journal of Northwest AF University(Natural Science)* **40** 217-23
- [22] Gong S G, Xie G L 2004 *Commands and Parametric Programming in ANSYS* Beijing: China Machine Press (in Chinese)
- [23] Zhang G X 2004 *Journal of Hydraulic Power (Beijing)* **56** 71-6 (in Chinese)
- [24] Zhu B F 1994 *Journal of Hydraulic Power (Beijing)* **46** 21-30 (in Chinese)
- [25] Zhu B F 2006 *Journal of Hydraulic Power (Beijing)* **37** 1424-32 (in Chinese)

Author



Hongyang Zhang, born on March 20, 1981, China

Current position, grades: North China University of Water Resources and Electric Power, Hydraulic Structure Engineering, Lecturer

Scientific interest: CAD modelling, model test, and Simulation analysis , Hydraulic Structure Engineering

Experience:

Date	School	Major	Degree
2000-2004,	Zhengzhou university,	Water resources and hydropower engineering,	Bachelor
2004-2007,	Zhengzhou university,	Hydraulic Structure Engineering ,	Master
2007-2010,	Hohai university,	Hydraulic Structure Engineering ,	Doctor
2010-now,	North China University of Water Resources and Electric Power,	Hydraulic Structure Engineering, Lecture	

Research on the operation modes of hydropower station based on complementary characteristics

Wenliang Yin^{1*}, Maoqing Xiang²

¹Chongqing Water Resources and Electric Engineering College/Changdu Prefecture Water Resources Bureau, Tibet

²Zhong County Water Supplies Bureau, Chongqing/Changdu Prefecture Water Resources Bureau, Tibet

Received 6 October 2013, www.tsi.lv

Abstract

Including wind-PV-ES (Wind/Photovoltaic/Energy storage) hybrid power generation system into the scheduling system of grid is the development tendency of safe grid-connection and operation of large wind-PV-ES hybrid power generation system. To solve the active power control problems in hybrid power generation system, this paper analyzes genetic algorithm and quantum genetic algorithm, and also analyzes the importance of energy storing devices in scheduling. Based on this, an optimization model of active power in wind-PV-ES is established. With the expectation of power output fluctuation of the power generation system as the objective function, the optimal scheduling scheme for the model is sought through genetic algorithm and quantum genetic algorithm respectively. The results of Matlab experiment show that the optimal scheduling scheme obtained by means of quantum genetic algorithm is superior to the scheduling scheme obtained by means of traditional genetic algorithm.

Keywords: Hybrid power generation system, Power control, Genetic algorithm, Quantum genetic algorithm

1 Introduction

With the developing of industry in modern times, energy problems have become a puzzle challenging the economic development of all countries in the world. As traditional energies are limited, for instance, petroleum and coal resources, in addition, traditional energies may bring unrecoverable damages to the environment, therefore, the development strategy of “Vigorously developing hydropower, optimally developing thermal power, actively developing nuclear power, and industriously to developing new energies” has been recognized by all countries in the world in recent years [1].

As a technically mature renewable energy power generation, hydropower is an important measure to guarantee energy supply. With the advantages of renewability, low operation expenses, being clean and environmental-friendly, strong ability in peak-load shaving and frequency modulation and the ability of restoring biological environment etc., more and more attention has been paid to hydropower. The construction of hydropower stations has taken shape in China. By far, over 170 hydropower stations have been built in 12 hydropower bases in upper and middle reaches of the Yellow River and Wujiang in China [2]. See Table 1 for the conditions of hydropower stations.

TABLE 1 Construction conditions of hydropower stations in China

Reservoir	Total storage (hundred million m ³)	Normal water level (m)	Installed capacity (ten thousand kw)	Annual energy output (hundred million kw)
The Three Gorges	393	175	1820	847
The Gezhou Dam	15.8	66	271.5	157
Ertan	58	1200	3300	170
Gongzui	3.1	48	70	34.18
Liyuan	7.27	1618	240	97.53
Ahai	8.82	1504	200	88.77
Guanyinyan	20.72	1134	300	122.4
Wudongde	76	975	870	387
Baihetan	188	820	1200	515
Xiluodu	126.7	600	1386	571.2
Xiangjiaba	51.63	380	640	307.47
Shuibuya	45.8	400	1600	39.2
Gaobazhou	4.3	80	25.2	8.98
Geheyan	34	200	120	30.4
Longyangxia	247	2600	128	23.6
Lijiaxia	16.5	2180	200	59
Liujiuxia	57	1735	122.5	55.8
Yanguoxia	2.2	1619	45.2	22.4
Wanjiazhai	8.96	977	108	27.5
Tianqiao	0.67	834	12.8	6.07
Liujiuxia	57	1735	122.5	55.8
Xiaolangdi	126.5	275	180	51
Xiaolangdi	126.5	275	180	51
...

See Figure 1 for the typical diagram of hydropower station.

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

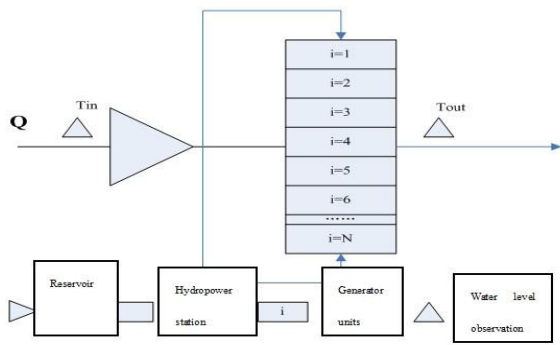
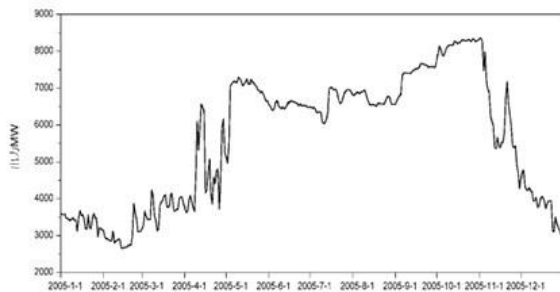


FIGURE 1 Typical diagram of hydropower station

The output power ability of a hydropower station is affected by the total storage, water level, installed capacity and the quantity of generator sets etc., therefore, the generated energy of a hydropower station is generally nonlinear. Figure 2 shows the power generating condition of a hydropower station in one year.



FIDURE 2 Diagram of the output of a hydropower station

Being affected by the said various factors, it is difficult for hydropower cater to the terminal demand during peak electricity demand period. Considering the actual problems, this paper mainly discusses multi-energy complementary power generation under the condition when hydropower cannot cater to the users' terminal demand under ideal status. At this moment, the output of the hydropower station can be deemed as a constant value. Among various renewable energies, solar energy and wind energy have been widely utilized, and they have been both used for power generation. However, both solar energy and wind energy are featured by low energy density and instability etc., therefore, using solar power generation or wind power generation alone cannot provide stable power supply output. Traditional processing method is to add energy storing devices additionally; however, the effects of such practice generally increases the equipment input of the power plants. In view of the complementary characteristics of the power supplies, the optimal scheduling scheme can be obtained through corresponding algorithms. The output of electric energy of wind-solar hybrid power generation can be perfected by scheduling the charging and discharging time of the energy storing devices every day, so as to enable wind-solar hybrid power generation to provide stable and reliable electric energy like traditional thermal power generation [3].

Each evolutionary algorithm has its own advantages and disadvantages. To make multiple intelligence algorithms complement each other's advantages, integrating different algorithms for intelligent optimization by following the thought of "optimal combination" is an important research direction [4]. In this paper, quantum genetic algorithm (QGA) and quantum-behaved particle swarm algorithm are adopted by integrating quantum genetic algorithm and quantum-behaved particle swarm algorithm, and comparing the results obtained by means of the new algorithm with the results obtained by means of traditional algorithms.

2 Introduction to the algorithms

Genetic algorithm is a random search algorithm deduced by simulating biological evolutionism (the principle of survival of the fittest in natural selection, i.e., selecting the superior and eliminating the inferior) in biology. It adopts a probabilistic method, which can quickly acquire the optimal space, and adjust the searching direction automatically without presetting the searching rules. This technology has been widely applied to sectors such as solving combination optimization solution, artificial intelligence, digital processing and machine learning etc. Quantum genetic algorithm is the product of the combination of quantum computation and genetic algorithm. This algorithm makes all the individuals in the space form an independent sub-swarm according to certain rules, and traverses each individual by coding; then it evolves each individual by means of quantum rotating gate method and dynamic adjusting rotation angle, with each individual subject to independent evolution, thus obtaining the optimal individual [5, 6].

2.1 GENETIC ALGORITHM

Genetic algorithm can be simply narrated as problems that the genes in organisms seek the most dynamic chromosome. Like the nature, for the answer designated by the configuration, what genetic algorithm needs to do is to obtain the daughter chromosome with better viability through genetic algorithm [7]. In this algorithm, firstly, some number codes (i.e. chromosome) of the target problem are produced randomly, and these number codes shall be defined as the first generation population. Fitness assessment shall be carried out on each individual with the preset fitness function, in which the individuals with poor fitness shall be weeded out and the individuals with good fitness shall be selected for offspring inheritance. The inherited individuals shall form a new population which becomes the second generation population, and the third, the fourth generation subject to inheritance.

Steps of algorithm:

Step 1: Initialization. Setting the size of the first generation population n , generally, n is between 30 and 60, and randomly producing an individual set P_i ($i=1,2,3...n$) as the first generation population; set the

maximum genetic algebra M .

Step 2: Selecting individuals. It is also known as individual assessment stage, i.e., carrying out fitness assessment on the individuals respectively with fitness function to select the individuals with good fitness and weed out those with poor fitness. Fitness function is f (also known as objective function), then $f(P_i)$ is the fitness of individual P_i .

$$e(P_i) = \frac{f(P_i)}{\sum (P_i)} \quad (1)$$

If P_i is selected, Formula 1 shall be used to calculate P_i 's inheritance frequency in the next generation. As known from (1), when $f(P_i)$ is large, its inheritance frequency in the next generation will be large; when $f(P_i)$ is small, its inheritance frequency in the next generation will be small; that is to say, the factor determining the inheritance frequency in the next generation lies in the fitness of the population where the individual is in.

Step 3: Crossover operation. This algorithm is similar to the biological hybridization in the nature, the process where the two parent generations conduct genic hybridization and recombination to produce the filial generation. Crossover operation is the core of genetic algorithm as well as the main way to produce the filial generation. Firstly, randomly select the same positions of two individuals in the parent generation to hybridize them in line with crossover probability. This method can be simply interpreted by the following individuals. There are two individuals $S1$ and $S2$, where $S1=111000$, $S2=000111$. According to the hybridizing rule, the two parent generations reciprocal interchange half of their information to recreate the filial generation $S1=000000$ and $S2=111111$.

Step 4: Mutation operation. According to the mutation principle of biology, this is the process during which mutation is conducted on local information of individuals in the parent generation with small probability event, and the mutant is passed onto the next generation through mutation operation.

For instance, for individual in parent generation $S=11111$, mutant of parent generation $S=100111$ can be obtained through mutation 2 and 3, and the mutant is passed on to the next generation as the new parent generation through cross operation.

A next generation population can be obtained through step 2, step 3 and step 4.

Step 5: Judgment termination condition. In case the fitness of the optimal individuals surpasses the threshold value given in advance, then inheritance shall stop. In case the fitness of the optimal individuals is still larger than the given threshold value, then the operations in step 2, step 3 and step 4 shall be continued, until meeting the maximum genetic algebra or being larger than the given threshold value.

2.2 QUANTUM GENETIC ALGORITHM

Step 1: Population initialization. According to formula

$$P_i = \begin{bmatrix} \cos(t_{i1}) & \cos(t_{i2}) & \cdots & \cos(t_{in}) \\ \sin(t_{i1}) & \sin(t_{i2}) & \cdots & \sin(t_{in}) \end{bmatrix} \quad (2)$$

The first generation population shall be formed by the n chromosomes produced randomly. Set the change value of the step length of initial rotation angle as θ , and set the probability of mutant chromosomes in each generation as mutation probability p_m .

Step 2: Solving spatial alternations. Map the approximate solution represented by each chromosome to continuous optimization problem equation by means of unit space $I_n=[-1,1]^n$

$$\begin{cases} \min f(x) = f(x_1, x_2, \dots, x_n) \\ \text{s.t. } a_i \leq x_i \leq b_i; i = 1, 2, \dots, n \end{cases} \quad (3)$$

Solve space Ω , and calculate the fitness of each chromosome according to formula

$$\text{fit}(x) = C_{\max} - f(x) \quad (4)$$

Record the optimal solution of the very generation as \hat{x}_0 , record the chromosome individual set of the very generation as \hat{p}_0 . Record the optimal solution of the previous generation as X_0 . Record the chromosome individual set of the previous generation as p_0 . If $\text{fit}(\hat{x}_0) > \text{fit}(x_0)$, then $p_0 = \hat{p}_0$; if $\text{fit}(\hat{x}_0) \leq \text{fit}(x_0)$, then $p_0 = \hat{p}_0$.

Step 3: for each quantum bit on each chromosome in the population, set the corresponding quantum bit in p_0 as the target, determine the size of the rotation angle in line with Formula (2) according to the rotation angle orientation, and update the quantum bit by means of the quantum rotating gate.

Step 4: carry out mutation on each chromosome in the population according to mutation probability by means of quantum negation gate.

Step 5: Return to Step 2 for circulative calculation until meeting the condition of convergence or the algebra reaches the maximum limitation [8, 9].

3 Researches on the optimal scheduling of Wind-PV-ES power generation system

The basic requirement of electrical power system is safe and reliable, economical and practical as well as superior quality of voltage. The requirement of economic

development on electrical power system increases year by year, meanwhile, electrical power system also sees earth-shaking changes year by year. Apart from traditional thermal power generation, a group of emerging operational modes of electrical power system are growing up gradually. Meanwhile, some new problems also occur accordingly [10]: environmental destruction problems are still unsolved, the users' electricity demand surpasses the transmission capacity of the grid, the users' requirement on the voltage quality of the grid is increasingly high, disturbance problem of the huge grid is increasing prominent, the maximum capacitance of the system turnaround cannot cater to the users' high load electricity demand, the users' technology for electric energy management lags behind etc.

At the beginning of this century, the US became the first to suffer from electricity shortage. The demand of users in some regions surpassed the generating capacity of power plants, which led to repeatedly power failure in those regions. After that, many countries in the world suffered from the problem of short supply of power plants to some different extents. China has seen power rationing phenomenon in many regions for many consecutive years since 2002, power shortage is particularly serious in those first-tier cities such as Beijing, Shanghai and Guangzhou etc. in summer. Electrical power system lags behind the users' electricity demand in respect of transmission capacity and system scheduling link. Moreover, this contradiction will continuously exert an influence worldwide in a long period, which will bring a long-term challenge for the operation of electrical power system [11].

Currently, electrical power system lacks of efficient compensation method and device for active power, while traditional method is to use standby generators, which has slow response speed. In electrical power system, in case of system failure, the standby generators cannot make corresponding change quickly enough, therefore, the stability of electrical power system cannot be guaranteed. Serious system failure may break down the electrical power system.

Energy storage technology is a kind of technology applied in electrical power system in early stage. This technology can solve the problem of unbalanced power supply of the grid to some extent. Currently, the main functions of energy storage technology in electrical power system include increasing the stability of electrical power system, improving power supply quality and voltage peak-load shaving etc.

The working principle of energy storing device is when the electrical load of the users are low, the energy storing device can charge as the load, and when the electrical load of the users are at the peak, the energy storing device can work as the power generating means [12]. This method can reduce the power consumption in electrical grid, and play the role of load shifting for the voltage in the grid, thus satisfy the stable work of the electrical power system. In addition, compared with

diesel generators, energy storing device is provided with the advantages of low electricity cost and fast response speed etc. [13].

4 Probability distribution of the output by wind power, photovoltaic power and energy storage system

4.1 PROBABILITY DISTRIBUTION OF THE OUTPUT BY WIND GENERATOR UNITS

The data in literature [14] indicates that the variation of wind speed with time can meet Rayleigh distribution:

$$f(v) = \frac{v}{\sigma_w^2} \exp\left(-\frac{v^2}{2\sigma_w^2}\right), \quad (5)$$

where $\sigma_w = \left(\frac{\pi}{2}\right)^{-\frac{1}{2}} v$, is the distribution parameter, v is the speed at a certain moment.

The generating power of wind generator units is as shown in Figure 3.

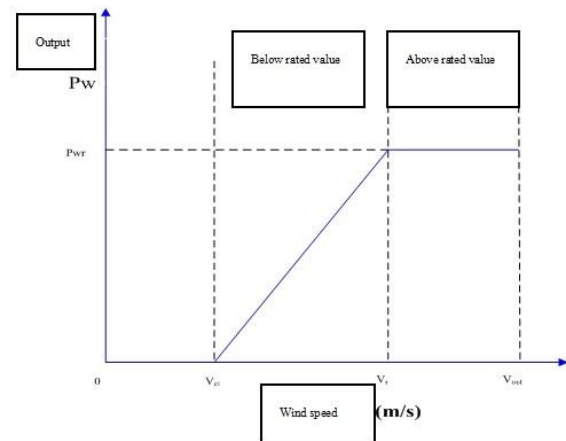


FIGURE 3 The generating power of wind generator units

The expression of the model that the wind generators units convert wind power is:

$$P_w = \begin{cases} 0 & 0 \leq v(t) \leq v_{ci} \text{ or } v(t) \geq v_{out} \\ av(t) + b & v_{ci} \leq v(t) \leq v_r \\ P_{wr} & v_r \leq v(t) \leq v_{out} \end{cases}, \quad (6)$$

where

$$a = \frac{P_{wr}}{v_r - v_{ci}}, \quad (7)$$

$$b = -av_{ci}. \quad (8)$$

In the expression: v_{ci} is for cut-in wind speed, v_r is

for rated wind speed, v_{out} is for cut-out wind speed, and P_{wr} is for the rated power.

The expectancy value of the output power of the wind generator units $E(P_w)$ and second order origin moment.

$$E(P_w) = a \left[-v_r \exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) + v_{ci} \exp\left(-\frac{v_{ci}^2}{2\sigma_w^2}\right) + \sqrt{2\pi}\sigma_w \left(\Phi\left(\frac{v_r}{\sigma_w}\right) - \Phi\left(\frac{v_{ci}}{\sigma_w}\right) \right) \right] - b \left[\exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) - \exp\left(-\frac{v_{ci}^2}{2\sigma_w^2}\right) \right] - P_{wr} \left[\exp\left(-\frac{v_{out}^2}{2\sigma_w^2}\right) - \exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) \right] \quad (9)$$

$$E(P_w^2) = 2a^2\sigma_w^2 \left[\left(\frac{v_{ci}^2}{2\sigma_w^2} + 1 \right) \exp\left(-\frac{v_{ci}^2}{2\sigma_w^2}\right) - \left(\frac{v_r^2}{2\sigma_w^2} + 1 \right) \exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) \right] + 2ab \left[-v_r \exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) + v_{in} \exp\left(-\frac{v_{ci}^2}{2\sigma_w^2}\right) + \sqrt{2\pi}\sigma_w \left(\Phi\left(\frac{v_r}{\sigma_w}\right) - \Phi\left(\frac{v_{ci}}{\sigma_w}\right) \right) \right] - b^2 \left[\exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) - \exp\left(-\frac{v_{ci}^2}{2\sigma_w^2}\right) \right] - P_{wr}^2 \left[\exp\left(-\frac{v_{out}^2}{2\sigma_w^2}\right) - \exp\left(-\frac{v_r^2}{2\sigma_w^2}\right) \right] \quad (10)$$

In the expression: Φ_0 is standard normal distribution function.

Therefore, the variance of the output power of the wind generator units can be easily obtained as follow:

$$D(P_w(t)) = E(P_w^2(t)) - E^2(P_w(t)) \quad (11)$$

where, $P_{s\max}(t)$ indicates the maximum power output every day, the maximum power output of the photovoltaic power generator units $P_{s\max}(t)$ is:

$$P_{s\max}(t) = r_{\max}(t)A\eta \quad (15)$$

4.2 PROBABILITY DISTRIBUTION OF THE DISPOSAL BY PHOTOVOLTAIC POWER SYSTEM

The experimental data in literature [16] indicates that the change of solar illumination intensity with time can meet Beta distribution:

$$f(r(t)) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{r(t)}{r_{\max}(t)} \right)^{\alpha-1} \left(1 - \frac{r(t)}{r_{\max}(t)} \right)^{\beta-1} \quad (12)$$

where $r(t)$ is for the solar illumination intensity at t ; $r_{\max}(t)$ is the maximum solar illumination intensity every day; Γ is Gamma function, α and β are shape parameters of Beta distribution.

Where the transient output power of photovoltaic power generator units at t is:

$$P_s(t) = r(t)A\eta \quad (13)$$

where, η is for photoelectric conversion efficiency (which is related to time point and the making technology of solar power generation panels); A is for the total area of solar power generation array [17].

The density function of photovoltaic output power is:

$$f(P_s(t)) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{P_s(t)}{P_{s\max}(t)} \right)^{\alpha-1} \left(1 - \frac{P_s(t)}{P_{s\max}(t)} \right)^{\beta-1} \quad (14)$$

To sum up, the expectation of the output power of photovoltaic power generator units $E(P_s(t))$ is:

$$E(P_s(t)) = \frac{\alpha}{\alpha+\beta} P_{s\max}(t) \quad (16)$$

The second order origin moment of the output power of photovoltaic power generator units $E(P_s^2(t))$ is:

$$E(P_s^2(t)) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} P_{s\max}^2(t) \quad (17)$$

Thus, the variance of the output power of photovoltaic power generator units $D(P_s(t))$ is:

$$D(P_s(t)) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} P_{s\max}^2(t) \quad (18)$$

5 Model of energy storing device

5.1 OBJECTIVE FUNCTION

While ignoring the power consumption in power grid transmission (or attribute this part of energy consumption to power consumption at client side), the users' service power of the electrical power system during t time interval is $P_L(t)$, the output power of the wind generator units is $P_w(t)$, the output power of photovoltaic power

generator units is $P_S(t)$, the sum of the powers in t time interval is recorded as power output fluctuation $P(t)$:

$$P(t) = P_L(t) - P_S(t) - P_W(t) - P_e(t), \quad (19)$$

when the system energy storing device charges, $P_e(t)$ will be a negative value; when the energy storing device discharges, $P_e(t)$ will be a positive value [19,20].

Assume N times of scheduling shall be carried out for the system, then the average value of power output fluctuation due to system scheduling is:

$$\min F(P(t)) = E \left[\frac{1}{T} \sum_{t=1}^T (P(t) - P_{avg})^2 \right] = E \left[\frac{1}{T} \sum_{t=1}^T P^2(t) - \left(\frac{1}{T} \sum_{t=1}^T P(t) \right)^2 \right] = \frac{1}{T} \sum_{t=1}^T E(P^2(t)) - \frac{1}{T^2} E^2 \left(\sum_{t=1}^T P(t) \right), \quad (21)$$

where

$$E(P^2(t)) = E^2(P_L(t) - P_S(t) - P_W(t) - P_e(t)) = (P_L(t) - P_e(t))^2 + E(P_W^2(t)) + E(P_S^2(t)) - 2(P_L(t) - P_e(t)) \left(E(P_S(t)) + E(P_W(t)) \right) + 2E(P_S(t))E(P_W(t)). \quad (22)$$

$$E^2 \left(\sum_{t=1}^T P_S(t) \right) = \sum_{t=1}^T D(P_S(t)) + \left(\sum_{t=1}^T E(P_S(t)) \right)^2, \quad (23)$$

$$E^2 \left(\sum_{t=1}^T P_W(t) \right) = \sum_{t=1}^T D(P_W(t)) + \left(\sum_{t=1}^T E(P_W(t)) \right)^2. \quad (24)$$

Substitute (22), (23), (24) in to expression (21), the objective function of the model can be obtained.

5.2 CONSTRAINT CONDITION

5.2.1 Installed capacity of the energy storing device

The electrical energy stored by the energy storing device cannot exceed the upper and lower limits of the capacity

$$EES_{\min} \leq EES(t) \leq EES_{\max}.$$

In the expression, $EES(t)$ is the electrical energy stored in the energy storing device at the end of t time interval; EES_{\max} and EES_{\min} are the upper and lower limits of energy storage [21].

5.2.2 Constraint of the charge-discharge power of energy storing device

The charge-discharge power of energy storing device must be less than the maximum discharge power of the energy storing device, i.e. $|PRES(t)| \leq PES_{\max}$.

In the expression, PES_{\max} is the maximum high charge and discharge power of the energy storing device.

5.2.3 Constraint of energy balance

$$EES(t) = EES(t-1) - PRES(t) \cdot \Delta t.$$

$$P_{avg} = \frac{1}{T} \sum_{t=1}^T P(t). \quad (20)$$

Through the above-mentioned analysis, the expectation of the power output fluctuation during the time interval of N times of scheduling everyday can be set as the objective function, and its function expression is:

In the expression, $EES(t)$, $EES(t-1)$ indicate the energy state of the energy storing device at the end of t and $(t-1)$ moment respectively; the difference between the two indicates the energy discharged or absorbed during t time interval; $PRES(t)$ stays unchanged in t time interval.

5.2.4 The energy in the energy storing device stays unchanged in the first and last time interval in one period

$$EES(t) = EES(0).$$

In the expression, $EES(0)$ indicates the initial energy state, $EES(t)$ indicates the energy state at T moment [19, 21].

6 Calculation examples and simulation results

Take an example of the wind-solar independent hybrid power generation system installed in a company domestically to optimize its energy storing. According to the above-mentioned design, the experimental data needed include: average sunshine data per hour every day at the installation site, see Table 2 for detailed data; the data of the average wind speed per hour in a day, see Table 3 for detailed data; the electricity utility condition at remote end, see Table 4 for the detailed data.

Parameters of wind generator units: cut-in wind speed v_{ci} is 4m/s, cut-out wind speed v_{out} is 24 m/s, rated wind speed v_r is 14 m/s, rated power P_{wr} is 200MW.

Parameters of photovoltaic power generation system: total area of solar array 3×10^6 m², photoelectric conversion efficiency η is 15%.

TABLE 2 Relative parameters of solar illumination intensity

Time interval	7	8	9	10	11	12	13	14	15	16	17	18	19
Maximum solar illumination intensity (W/m ²)	3	15	41	53	68	86	89	79	74	58	46	12	2

TABLE 3 Parameters of wind speed change

Time (h)	1	2	3	4	5	6	7	8	9	10	11	12
Wind speed (m/s)	20	15	14	8	10	14	15	16	16	7	8	10
Time (h)	13	14	15	16	17	18	19	20	21	22	23	24
Wind speed (m/s)	9	18	20	16	18	17	15	18	16	18	21	22

TABLE 4 Load parameters

Time (h)	1	2	3	4	5	6	7	8	9	10	11	12
Load (MW)	925	880	800	715	710	825	960	1075	1100	1125	1165	1225
Time (h)	13	14	15	16	17	18	19	20	21	22	23	24
Load (MW)	1200	1050	975	960	950	1025	1150	1150	1215	1150	970	935

6.1 RESULTS OF GENETIC ALGORITHM OPTIMIZATION

Parameters of genetic algorithm: take the population size as 200, crossover probability as 0.8, mutation probability as 0.1, evolution algebra as the 500th algebra, the variable number as 24, which represent the charge and discharge condition of the energy storing device per hour in the 24 hours every day. According to the said model, the following changing condition of objective function with the evolution algebra shall be obtained, which is as shown in Figure 4. The optimal charge and discharge time is as shown in Figure 5.

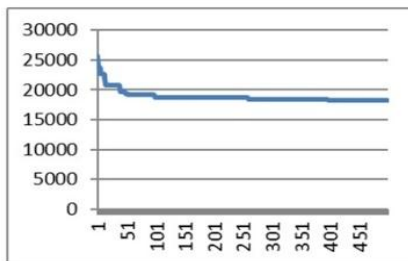


FIGURE 4 The changing condition of objective function value obtained by means of genetic algorithm

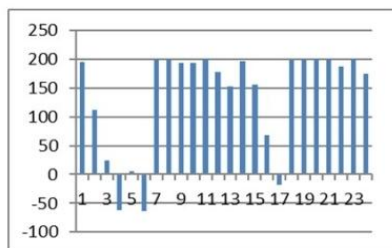


FIGURE 5 The optimal charge and discharge time obtained by means of genetic algorithm optimization

The final objective function value is converged to 1.8305×10^4 .

6.2 RESULTS OF QUANTUM GENETIC ALGORITHM OPTIMIZATION

Parameters of quantum genetic algorithm: take the

population size as 200, rotation angle length as 0.001π , evolution algebra as the 500th algebra, the variable number as 24, which represent the charge and discharge condition of the energy storing device per hour in the 24 hours every day. According to the previous discussion, the following changing condition of objective function with the evolution algebra shall be obtained, which is as shown in Figure 6. The optimal charge and discharge time is as shown in Figure 7.

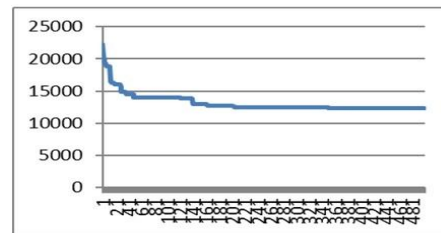


FIGURE 6 The changing condition of objective function of by means of quantum genetic algorithm

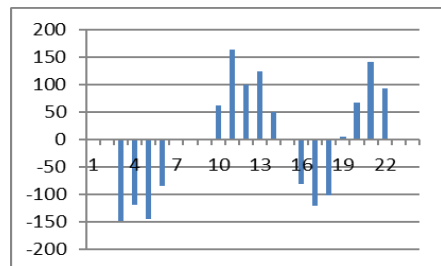


FIGURE 7 The optimal charge and discharge time value obtained by means of quantum genetic algorithm

The final objective function value is converged to 1.2424×10^4 .

The results show that, reasonably arranging the charge and discharge time of the energy storing device can minimize the fluctuation of the system, thus making the system more stable. In the model established in this paper, it's obvious that quantum genetic algorithm is superior to genetic algorithm, with the optimization results obtained by quantum genetic algorithm more reasonable. This indicates that quantum genetic algorithm has broad application prospect and efficient optimization efficiency.

7 Conclusions and prospect

The model constructed in this paper takes into consideration the complementary characteristics among the power generations systems, and obtains the optimal scheduling scheme by means of genetic algorithm and optimized genetic algorithm. Reasonable charge and discharge for the energy storing device can effectively improve and even solve the deficiencies in respect of power generation by renewable resources, which can effectively improve the stability of the system's output voltage. In the model constructed in this paper, the assumption is wind speed subject to Rayleigh distribution and solar illumination intensity subject to Beta

distribution, the objective function is power output fluctuation, corresponding constraint conditions are set, and the optimal scheduling scheme is obtained by means of genetic algorithm and quantum genetic algorithm. The experimental results show that the optimal scheduling scheme obtained by means of quantum genetic algorithm is superior to that obtained by means of traditional genetic algorithm. The quantum genetic algorithm adopted in this paper is also a relatively new optimization algorithm whose application field is very small; however, seen from some respects in this paper, the effectiveness of quantum optimization algorithm can completely surpass traditional algorithm, and can be widely applied.

References

- [1] Dong Yingrui, Yang Jinming, Hu Haiping 2012 Coordination Control of Distributed Wind-solar Hybrid System *Electrical Measurement & Instrumentation* **49**(554) 48-51
- [2] Xiong Liang, Zou Xuan 2010 Research on the Control Method of Direct-drive Wind Power System *Electrical Measurement & Instrumentation* **47**(530) 31-4
- [3] John H, David C Y, Kalu B 2008 An economic scheduling model incorporating wind power *IEEE Transactions on Energy Conversion* **23**(2) 603-11
- [4] Yang Hong-xing, Lu Lin, Zhou Wei 2007 A Novel Optimization Sizing Model For Hybrid Solar-Wind Power Generation System *Solar Energy* **81**(1) 76-84
- [5] Kellogg W D, Nehrir M H, Venkataramanan G, Gerez V 1998 Generation Unit Sizing and Cost Analysis For Stand Alone Wind, Photovoltaic, and Hybrid Wind/PV Systems *IEEE Transactions on Energy Conversion* **13**(1) 70-5
- [6] Liang R H, Liao Jian-hao 2007 A fuzzy optimization approach for generation scheduling with wind and solar energy systems *IEEE Transactions on Power Systems* **22**(4) 1665-74
- [7] Farhat I A, Hawary M E 2010 Dynamic adaptive bacterial foraging algorithm for optimum economic scheduling with valve-point effects and wind power *IET Generation, Transmission & Distribution* **4**(9) 989-99
- [8] Bouffard F, Galiana F D 2008 Stochastic security for operations planning with significant wind power generation *IEEE Trans on Power Systems* **23**(2) 306-16
- [9] Meibom P, Barth R, Hasche B 2011 Stochastic optimization model to study the operational impacts of high wind penetrations in Ireland *IEEE on Power Systems* **26**(3) 1367-79
- [10] Botterud A, Zhou Z, Wang J 2011 Unit Commitment and Operating Reserves with Probabilistic Wind Power Forecasts *Proceedings IEEE Trondheim Power Tech 2011* Trondheim, Norway 520-7
- [11] Liu Xian, Xu Wei-sun 2010 Economic Load Scheduling Constrained by Wind Power Availability: A Here-and-Now Approach *IEEE Transactions on Sustainable Energy* **1**(1) 2-9
- [12] Wang Haohuai, Tang Yong, Hou Junxian 2012 Equivalent Method of Integrated Power Generation System of Wind, Photovoltaic and Energy Storage in Power Flow Calculation and Transient Simulation *Proceedings of the CSEE* **32**(1) 1-8
- [13] Xu Lin, Ruan Xinbo, Zhang Buhan 2012 An Improved Optimal Sizing Method for Wind-solar-battery Hybrid Power System *Proceedings of The CSEE* **32**(25) 88-98
- [14] Sifuentes S W, Vargas A 2007 Hydrothermal scheduling using benders decomposition: accelerating techniques *IEEE Trans on Power Systems* **22**(3) 1351-9
- [15] Bisanovic S, Hajro M 2008 Hydrothermal self-scheduling problem in a day-ahead electricity market *Electric Power System Research* **78**(9) 1579-96
- [16] Wu Hong-yu, Guan Xiao-hong, Zhai Qiao-zhu 2009 Short-term hydrothermal scheduling using mixed-integer linear programming *Proceedings of the CSEE* **29**(28) 82-8
- [17] Horsley A, Wrobel A J 2007 Profit-maximizing operation and valuation of hydroelectric plant: A new solution to the Koopmans problem *Journal of Economic Dynamics and Control* **31**(3) 938-70
- [18] Zhu Jian-quan, Wu Jie-kang 2008 An uncertainty model for short-term optimization scheduling of hydrothermal power system *Automation of Electric Power Systems* **32**(6) 51-4
- [19] Basu M 2004 An interactive fuzzy satisfying method based on evolutionary programming technique for multiobjective short-term hydrothermal scheduling *Electric Power Systems Research* **69**(2) 277-85
- [20] Shao Bao-zhu, Wang You-yin, Song Dan 2010 Application of electricity futures in optimized allocation of thermal and hydroelectric power resources *Power System Technology* **34**(11) 170-5
- [21] Ali Tofighi 2013 Performance Evaluation of PV Module by Dynamic Thermal Model *Journal of Power Technologies* **93**(2) 111-21

Authors



Yin Wenliang, born in November, 1983, Yongchuan District, Chongqing city, P.R. China

Current position, grades: lecturer China of Chongqing Water Resources and Electric Engineering College.

University studies: B.Sc. of Engineering in Hydrology and Water Resources Engineering from Guizhou University in China. M.Sc. from Guizhou University in China.

Scientific interest: Hydrology, water resources, hydropower planning.

Publications: more than 10 papers published in various journals.

Experience: teaching experience of 8 years, 5 scientific research projects.



Xiang Maoqing, born in May, 1971, Zhong County, Chongqing city, P.R. China

Current position, grades: Professor level senior engineer China of Zhong County Water Supplies Bureau.

University studies: graduated from the water conservancy and Hydropower Engineering Sichuan Province Wanxian water conservancy and electric power professional school in 1992 in China.

Scientific interest: Water conservancy and hydropower planning, water conservancy and hydropower construction project.

Publications: more than 30 papers published in various journals.

Experience: 18 years of experience in water conservancy and Hydropower engineering, 15 scientific research projects.

Finite element analysis of the dynamic response of the cardiovascular system to the blunt ballistic impact

Hai Liu¹, Jing Chen^{1*}, Jian-yi Kang¹, Xiao-xia Li¹, Ivan Azhari²

¹ State Key Laboratory of Trauma, Burns, and Combined Injury, Institute of Surgery Research, Daping Hospital, Third Military Medical University, Chongqing 400042, China

² ACM Medical Health Equipment Design & Consultant Center, San Francisco, CA, USA

Received 6 May 2014, www.tsi.lv

Abstract

On the basis of the Chinese Visible Human Dataset (CVHD), a three-dimensional human finite element model that includes skin, muscle, bone, the lungs, the heart and the vascular trunk was developed. In the LS-DYNA software environment, a numerical simulation of the blunt ballistic impact, which was caused by a 5.56-mm rifle bullet moving with the speed of 910 m/s toward a human torso wearing a composite body armor vest, was performed, and the stress and pressure response of the cardiovascular system were calculated. The simulation results demonstrated that the blunt ballistic impact introduced a high-frequency pressure response on the chambers of heart, which was characterized by a high amplitude and short duration. The peak values of the pressure waves, measured at the ascending aorta and superior vena cava ports, were 659.3 kPa and 542.8 kPa respectively, which suggested that the blunt ballistic impact on the chest would result in injury to distant target organs through the cardiovascular system. The computational results of this model can provide a basis for predictions of heart injuries, in-depth studies of the mechanical mechanism of cardiovascular injuries to blunt ballistic impacts and further improvements in protective equipment.

Key words: blunt ballistic impact, cardiovascular system, dynamic response, finite element analysis

1 Introduction

In regional conflicts and violent terrorist attacks, gunshot wounds are one of the main threats. Use of body armor can effectively reduce the occurrence of penetrating wounds, but the blunt ballistic impact generated from body armor after it is hit by a bullet can still possibly cause damage to human tissues and organs [1]. The cardiovascular system is often an important target of blunt impacts. In an accident investigation in 2001, Siegel discovered that whether cardiac trauma occurred after the blunt impact to the chest was an important determinant of survival [2]. The cardiovascular system injuries caused by the blunt ballistic impact are very frequent in various types of military activities, but the injury mechanism, especially as to the biomechanical mechanism, is still unclear. Currently, physical models [3], animal experiments [4, 5], corpse tests [6, 7] and digital simulation models [8,9] are used for studying the dynamic response of blunt impact to the chest. Some researchers have performed finite element analyses of aortic rupture caused by blunt impacts [8]. However, the heart is treated as a single homogeneous medium in most studies, and this treatment is not sufficient for calculating the actual dynamic response of the cardiovascular system to the blunt ballistic impact. Until now, a general finite element simulation of the cardiovascular system's response to the blunt ballistic impact has not been reported. In this paper, a human torso finite element

model, based on CVHD, to perform the numerical simulation of the dynamic response to the impact of a 5.56-mm rifle bullet on body armor was constructed. Through the analysis of the stress and pressure distribution rules of cardiovascular system, the characteristics and mechanisms of cardiovascular injuries caused by the blunt ballistic impact would be discussed. The research can provide in-depth studies of the mechanical mechanism of the cardiovascular injuries caused by the blunt ballistic impact and further improvements in protective equipment.

2 Principles of cardiovascular system's response to the blunt ballistic impact

The response of cardiovascular system to the blunt ballistic impact is a complex mechanical problem related to the processes such as how bullets penetrate body armor, the collision between the armor and the human body, and the propagation of pressure waves in the human body.

When bullets penetrate body armor, if the contact stress σ is greater than dynamic yield limit of σ_r^D , plastic deformation of the bullets or body armor occurs. According to the basic formula of stress waves, the backward speeds after the bullets and body armor impact v_1, v_2 are shown as follows respectively,

*Corresponding author e-mail: chenjing9811@dphospital.tmmu.edu.cn

$$v_1 = \frac{\sigma}{\rho_p c_{ep}}, \quad (1)$$

$$v_2 = \frac{\sigma}{\rho_t c_{et}}. \quad (2)$$

At this time, the relative speed VEA is the bullet-impact speed limit, which is also known as H-K speed limit:

$$v_{EA} = \sigma_r^D \left(\frac{1}{\rho_p c_{ep}} + \frac{1}{\rho_t c_{et}} \right), \quad (3)$$

where ρ_p is the density of the bullet material, c_{ep} is the elastic wave velocity of the bullet material, ρ_t is the density of the protective material, and c_{et} is the elastic wave velocity of the protective material.

When the relative speed is increased, the relative compressibility of the solid is reduced. Consequently, a pressure wave forms in the solid. The relative speed at this time, VHA, is known as the speed limit of fluid deformation.

$$v_{HA} = \sqrt{\frac{K_t}{\rho_t}}, \quad (4)$$

where K_t is the volume compression modulus of the protective material.

After instantaneous deformation of the body armor, a collision with the body surface occurs, and the remaining energy is transmitted to the body in the form of pressure wave. According to the classical theory of biomechanics, the vast majority of human tissues are nonlinear viscoelastic materials, and their response depends on the loading conditions. According to Ogden's [10] nonlinear elastic material theory and Christensen's [11] viscoelasticity theory, the stress-strain relationship is as follows:

$$\sigma = \mu * \lambda_x^{\alpha-1} + \int_0^t G * e^{-\beta(1-\tau)} \frac{\partial \varepsilon}{\partial \tau} d\tau, \quad (5)$$

where α is the Ogden model material parameter, μ is super elastic modulus of human tissue, β is the Christensen model parameter, and G is viscoelastic modulus of human tissue.

3 The mechanical mechanism of cardiovascular injuries to the blunt ballistic impact

Two mechanisms are suspected to explain cardiovascular injuries to the blunt ballistic impact: First, compression and shear force of the sternum and ribs acts on the

intermediate organs, which causes structural damage to the heart and surrounding tissue; Second, transmission of pressure wave in the cardiovascular system causes damage to the distant parts of the body. The cardiovascular system is a pressure circulation loop, and the blood is an incompressible continuous fluid medium; therefore, pressure wave attenuation is less in the cardiovascular system than in other tissue. The transmission of pressure wave in blood vessels may cause injury to distant target organs (such as the brain). Such injury effects are called remote effects. Bir confirmed that the blunt ballistic impact can cause structural damage to the heart, lungs and other organs [7]. An animal study on behind-armor blunt trauma demonstrated that exposed animals exhibited decreased cardiac capacity [12]. Cripps et al claimed that the in vivo pressure wave that are generated by the blunt ballistic impacts are the main cause of damage to internal organs [13]. Cernak found that the pressure wave caused by explosion are transmitted to the distant parts through the cardiovascular system [14]. Courtney argued that the blunt ballistic impact may cause similar injuries to the human body [15]. These results suggest that the increase in blood vessels pressure during the blunt ballistic impact may be significant enough to cause the cardiovascular injuries.

4 Numerical simulation

The dataset was obtained from CVHD provided by the Digital Medicine Institute of the Third Military Medical University at <http://cvh.tmmu.edu.cn/cvhstore/index.asp>. This resource was freely available. The original specimen for the digitized human dataset was the dead body of a 35-year-old Chinese male with a height of 170 cm and weight of 65 kg. After frozen embedding, milling was performed using an industrial milling drill, and the cross-sections were then photographed. The dataset included 2518 cross-sectional images with a section spacing of 1.0 mm and a horizontal pixel size of 0.167 mm. The continuous images of the layers 1405-1709 were selected for reconstruction of a three-dimensional model of the human chest.

The human images were imported into the Mimics version 16.0 software. In the segmentation module, three-dimensional geometric model was established. The generated geometrical model was saved in STL (standard triangle language) files. The STL format of the geometric model was imported to HyperHesh version 10.0, which was used to generate the corresponding shell and solid elements. The chest model was composed of 15 parts and included 91,339 nodes, 527,020 tetrahedral elements and 9,612 shell elements. The skin was constructed from shell elements, whereas the muscle equivalents, bone, the lungs, the heart, blood vessels and blood were constructed from solid elements. The complete chest finite element model is shown in Figure 1. The cardiovascular system includes cardiac muscle, the left ventricle, the left atrium, the right ventricle, the right atrium, the ascending aorta, the superior vena cava, the

inferior vena cava and blood. The details of the cardiovascular model are shown in FIGURE 2. Linear elastic properties were used for bone and viscoelastic properties were used for other tissues and organs. The material parameters and constitutive relations are described in the references [8, 16, 17, 18]. The material parameters of various tissues are presented in Table 1.

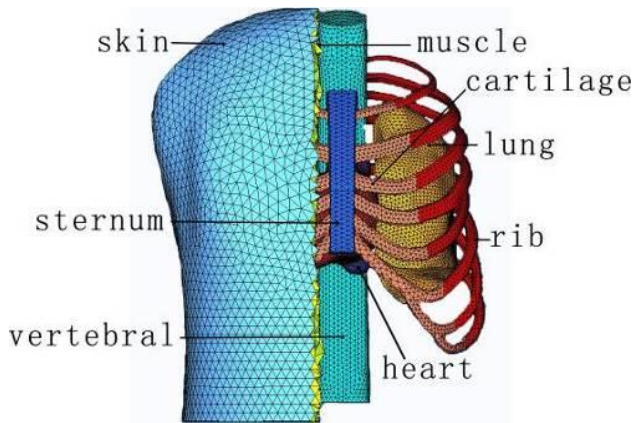


FIGURE 1 Complete finite element model of the chest

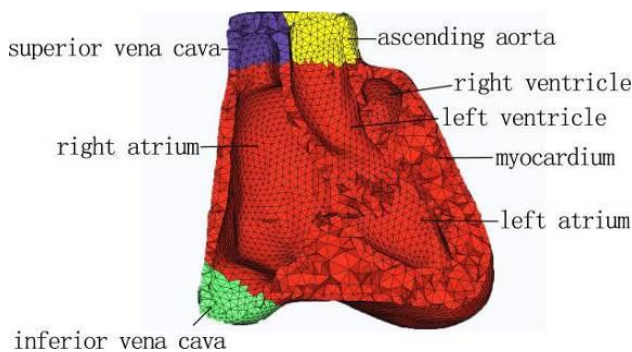


FIGURE 2 Finite element model of the cardiovascular system

A 5.56-mm rifle bullet, which was composed of the bullet core and the shell case, was used. The MAT_JOHNSON_COOK elastic-plastic material model in the GRUNEISEN state equation was adopted. The body armor was a composite structure that was composed of a panel of alumina ceramic and a backboard of high-density polyethylene fibre. For the ceramic material, the MAT_JOHNSON_HOLMQUIST-CERAMICS damage material model was used. For the high-density polyethylene fibre plate, the MAT_COMPOSITE_DAMAGE composite material model was used. A face-intrusive contact between the bullet and body armor was used to simulate the penetration of a bullet into the body armor. The bullet impact point was located in the middle of the sternum, and the bullet's initial speed was 910 m/s. The bullet was shot along the horizontal direction. A detailed model chart is presented in FIGURE 3. After a K file was generated in HyperMesh, it was imported to the finite element software package ANSYS/LS-DYNA version 11.0 to perform the computation.

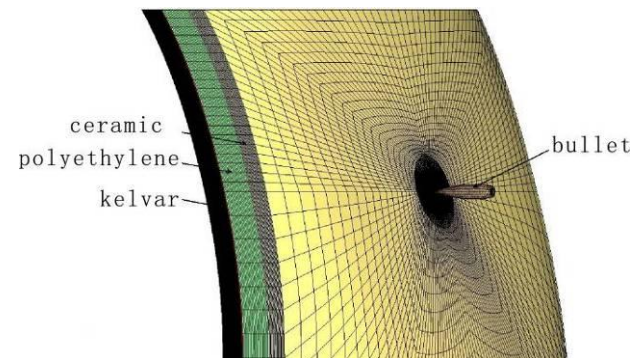


FIGURE 3 The finite element model of bullets and composite body armor

TABLE 1 The material parameters of the various tissues in the human finite element model

Tissues	ρ (kg/m ³)	K/GPa	G_0 /KPa	G_∞ /KPa	β	E/GPa	ν
Skin	1200	2.9	200	195	0.1		
Muscle equivalents	1120	1.03	200	195	0.1		
Sternum	1250					9.5	0.25
Cartilage	1170					0.0025	0.4
Ribs	1180					9.5	0.2
Spine	1330					0.355	0.26
Cardiac muscle	1120	0.744	67	65	0.1		
Lung	600	0.744	67	65	0.1		
Artery	1120	0.744	67	65	0.1		
Blood	1060			2000		1×10^{-10}	0.5

Note: β - attenuation coefficient; G_0 -short-term elastic shear modulus; G_∞ - long-term elastic shear modulus; K - elastic bulk modulus; E - modulus of elasticity; ν - Poisson's ratio

5 Results and analysis of numerical simulation

After the LS-DYNA computation was finished, the post-processing program prepost was used to determine the stress distribution at different times and the pressure response at different locations.

When the bullet hit the body armor, the stress was transmitted to the heart and other internal organs in the form of wave. At $t=48 \mu s$, the stress wave reached the

surface of the heart. Over time, the stress fields expanded to the surroundings. At $t = 2000 \mu s$, the stress wave covered all areas of the heart surface and transferred to the inferior vena cava. FIGURE 4 shows the Von Mises equivalent stress distribution on the surface of the heart at different times. Figure 5 presents the Von Mises equivalent stress distribution of cross-section of the heart, which shows the distribution of the stress wave in the chambers of the heart at different times.

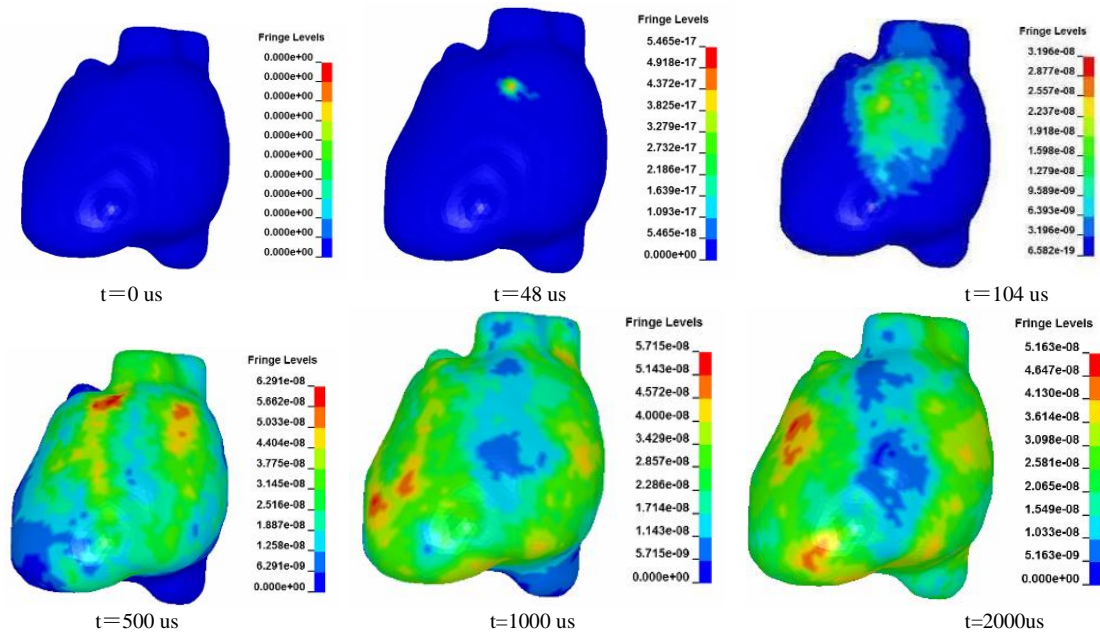


FIGURE 4 Equivalent stress distribution on the surface of the heart at different times

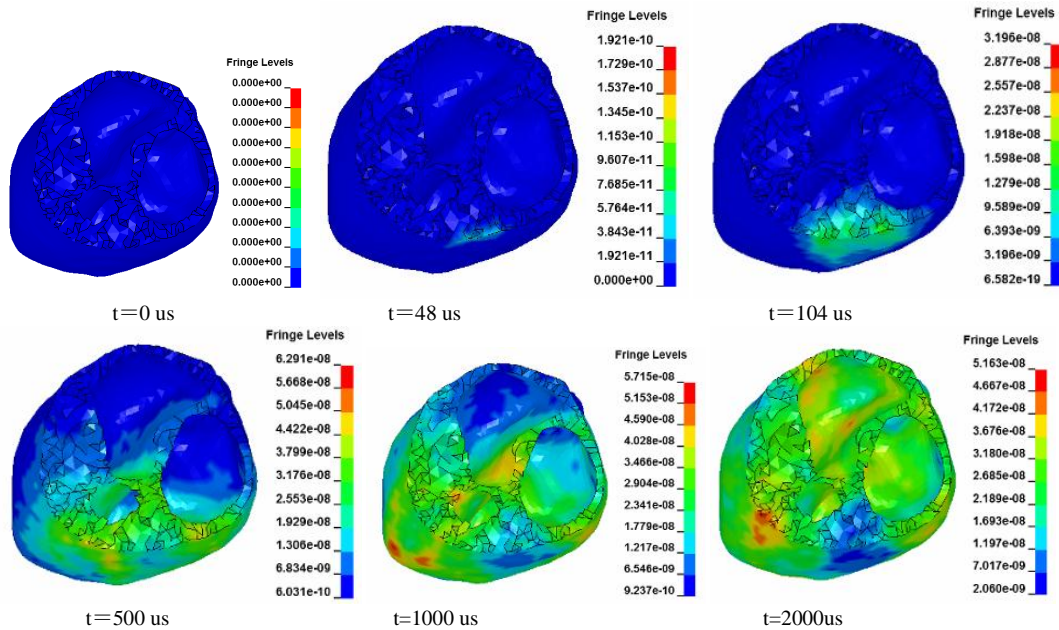


FIGURE 5 Equivalent stress distribution in the chambers of the heart at different times

The blunt ballistic impact introduced a high-frequency pressure response in the human cardiovascular system, which was characterized by a high amplitude and short duration. The pressure curves for the ventricles and atria are shown in Figure 6. The peak of the instantaneous pressure in the left ventricle was the greatest, with a value

of approximately 0.8012 MPa, and the peak of the instantaneous pressure in the left atrium was the lowest, with a value of approximately 0.3167 MPa. Figure 7 shows the pressure curves in the ascending aorta and superior vena cava.

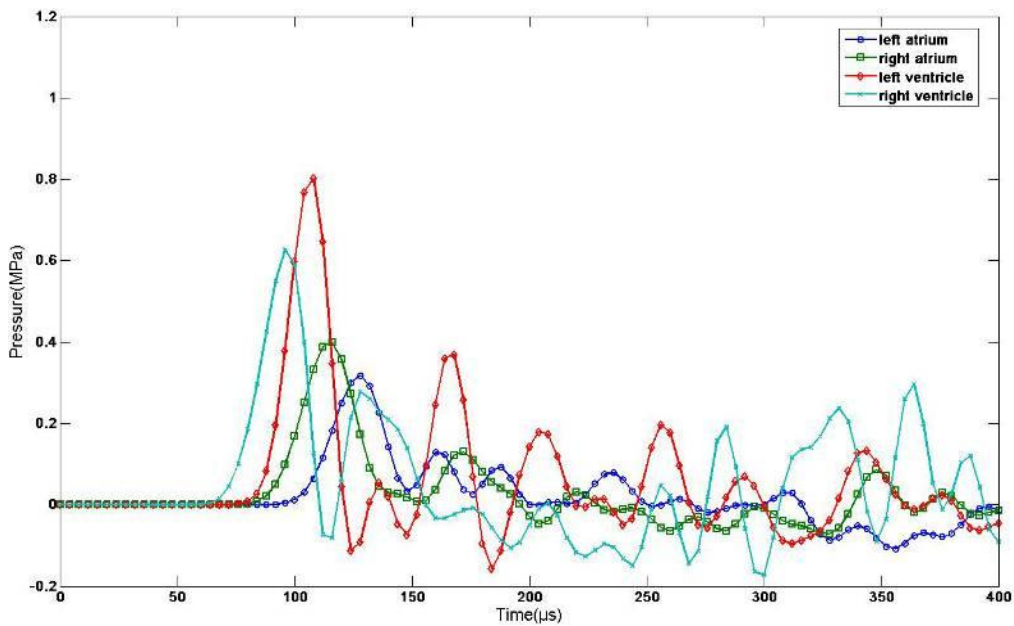


FIGURE 6 The pressure curves for the atria and ventricles

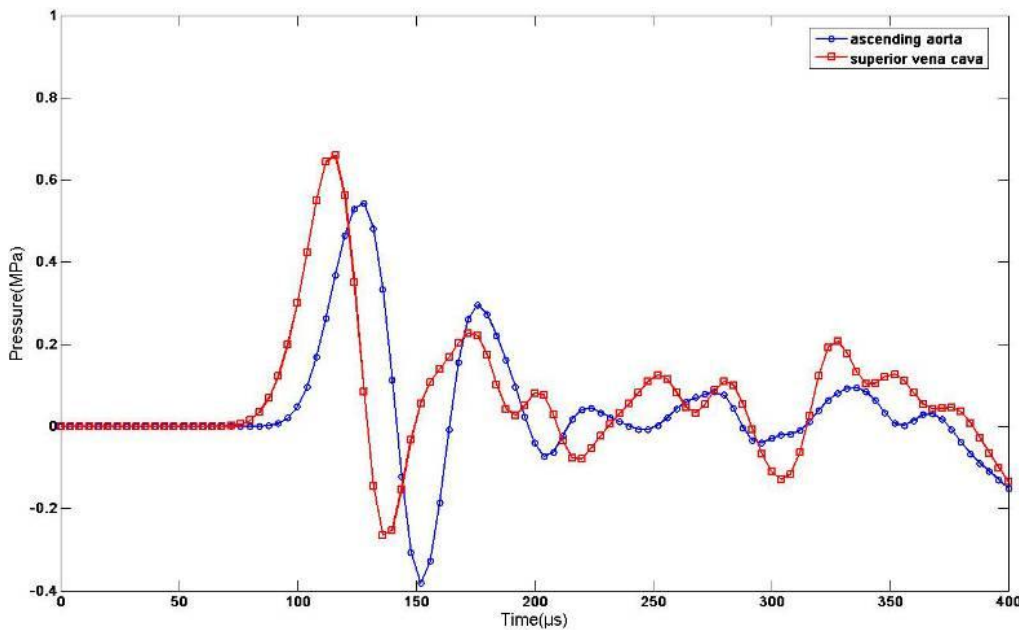


FIGURE 7 The pressure curves for the ascending aorta and superior vena cava

In this model, the maximum Von Mises stress of the heart appeared at the left ventricle, and the sites of the peak values of pressure waves were the left ventricle, right ventricle, right atrium, and left atrium, in descending order; the pressure peak in the ventricle was 2-3 times that in the atria. The results suggested that the ventricle is easier to rupture than the atria, which was consistent with the results of an accident investigation from Siegel [2]. When the intracranial pressure reaches 100-300 kPa, it can cause mild-to-moderate brain injuries [19]. The peak values of high-frequency pressure waves measured in the ascending aorta and superior vena cava port were 659.3 kPa and 542.8 kPa respectively, that suggested the pressure wave generated from the blunt ballistic impact might be transmitted from the

cardiovascular system to the brain. Currently, the rules of transmission of high-frequency pressure wave in the cardiovascular system are still unclear, and they should be further studied in the future.

6 Conclusions

The present study proposes a computer-based simulation approach to investigate the response of cardiovascular system to blunt ballistic impacts. A three-dimensional finite element model, which includes human cardiovascular organs is developed. By using this model, the numerical computation of the stress distribution and pressure response of cardiovascular system under the blunt ballistic impact are implemented. The human injury



results predicted by the model are found to agree with the injury effects observed in animal experiments and reported in the literatures. The model accurately reflects

the morphological structure of the heart cardiovascular system, that will be helpful in the design of impact protection systems.

References

- [1] Cannon L 2001 *Journal of the Royal Army Medical Corps* **147**(1) 87-96
- [2] Siegel J H, Smith J A, Tenenbaum N 2001 *Annual proceedings/Association for the Advancement of Automotive Medicine* **46** 315-38
- [3] Liu H, Kang J, Chen J 2012 *International journal of medical sciences* **9**(8) 655
- [4] Kunz S N, Arborelius U P, Gryth D 2011 *Journal of Trauma and Acute Care Surgery* **71**(5) 1134-43
- [5] Zhang, B, Huang Y, Su Z 2011 *Journal of Trauma and Acute Care Surgery* **71**(6) 1680-8
- [6] Baqué P, Serre T, Cheynel N 2006 *Journal of Trauma and Acute Care Surgery* **61**(3) 586-91
- [7] Bir C, Viano D, King A 2004 *Journal of biomechanics* **37**(1) 73-9
- [8] Richens D, Field M, Hashim S 2004 *European journal of cardio-thoracic surgery* **25**(6) 1039-47
- [9] Roberts J C, Merkle A C, Biermann P J 2007 *Journal of Biomechanics* **40**(1) 125-36
- [10] Ogden R W 1972 *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **326**(1567) 565-84
- [11] Christensen R 1982 *Theory of viscoelasticity: an introduction* Elsevier p 26
- [12] Drobin D, Gryth D, Persson J K 2007 *Journal of Trauma-Injury, Infection, and Critical Care* **63**(2) 405-13
- [13] Cripps N P J, Cooper G J 1996 *Journal of Trauma-Injury, Infection, and Critical Care* **40**(3S) 206S-11S
- [14] Cernak I, Wang Z, Jiang J 2001 *Journal of Trauma-Injury, Infection, and Critical Care* **50**(4) 695-706
- [15] Courtney A C, Courtney M W 2009 *Medical Hypotheses* **72**(1) 76-83
- [16] Kang J, Chen J, Dong P 2012 A Computer-Based Algorithm for Mechanical Response of Chest Wall to Blunt Ballistic Impact *International Review on Computers & Software* **7**(7)
- [17] Saraf H, Ramesh K T, Lennon A M 2007 *Journal of biomechanics* **40**(9) 1960-7
- [18] Caruso K S, Hijuelos J C, Peck G E 2006 *Journal of advanced materials* **38**(3) 27-36
- [19] Toth Z, Hollrigel G S, Gorcs T 1997 *The Journal of neuroscience* **17**(21) 8106-17

Authors

	<p>Hai Liu, born in July, 1984 ,Chongqing, P R China</p> <p>Current position, grades: researcher assistant at Research Institute of Surgery in Third Military Medical University, China. University studies: graduate student of Third Military Medical University University in China Scientific interest: Biomechanics and Numerical Simulation</p>
	<p>Jing Chen, born in January, 1968, Chongqing, P R China</p> <p>Current position, grades: researcher at Research Institute of Surgery in Third Military Medical University, China. University studies: PhD degree in Biomedical Engineering from Chongqing University in China Scientific interest: wound ballistics and biomechanics mechanism of trauma caused by gunshot and blast wave</p>
	<p>Jian-yi Kang, born in September, 1973, Chongqing, P R China</p> <p>Current position, grades: researcher at Research Institute of Surgery in Third Military Medical University, China. University studies: PhD in Biomedical Engineering from Chongqing University in China Scientific interest: wound ballistics and Numerical Simulation of human injury caused by gunshot and blast wave</p>
	<p>Xiao-xia Li, born in January, 1976, Chongqing, P R China</p> <p>Current position, grades: researcher assistant at Research Institute of Surgery in Third Military Medical University, China. University studies: BSc in mechatronic engineering from University of Jinan in China Scientific interest: Biocomputing and Biomechanics</p>

Calculation of helicopter maneuverability in forward flight based on energy method

Nanjian Zhuang, Jinwu Xiang, Zhangping Luo*, Yiru Ren

School of Aeronautic Science and Engineering, Beihang University, Haidian Dist, Beijing, China

Received 1 March 2014, www.tsi.lv

Abstract

A new method for calculating helicopter maneuverability in forward flight is proposed. Empirical equations for evaluating rotor required power are employed. Using energy method, an algorithm to calculate the available overloads, rate of climb and flight trajectory is given. The maneuver performance of AH-1G helicopter is investigated and three kinds of maneuvers including level acceleration, deceleration turning and turning climb followed by accelerating climb are calculated and analysed. Numerical results indicate that the method is effective and feasible, even for three dimensional maneuvering problems. In addition, the method can be applied to predict flight trajectory during forward flight.

Keywords: Helicopter, Maneuverability, Energy Method

1 Introduction

With the development of aviation technology, plenty of new roles and missions are assigned for helicopter. Sometimes helicopters were required for rescues and aid in an unknown, dynamic and potentially hostile environment. In order to finish the specified task successfully, helicopters should equip collision avoidance system, which demands real-time and accurate helicopter maneuverability and flight trajectories [1].

Due to the complex helicopter aerodynamics, the maneuvering process need much more time based on the existing method. To obtain the designed performance, energy method is used by a lot of researchers to investigate two-dimensional (2-D) helicopter maneuver flight problems. Basic studies about helicopter required power were conducted in detail [2-4]. Helicopter performances were investigated using energy balance method [5, 6]. Some maneuver characteristics were studied by Xu [7] and Mikhailov [8-10]. Aerobatic maneuvers were analysed in detail based on mathematical description by Cao [11] and Hu [12, 13].

However, there are few researches on helicopter 3-D maneuver using energy method. In this paper, an algorithm based on energy method is described. Using the data of AH-1G helicopter, three kinds of maneuvers are analysed, including level acceleration, decelerating turn, turning climb followed by accelerating climb.

2 Mathematical model

The energy state of a helicopter can be written as:

$$E = \frac{1}{2} mV^2 + mgh + \frac{1}{2} I\Omega^2, \quad (1)$$

where m is mass of helicopter, I is total rotor inertia, Ω is rotor rotational speed. By taking the partial derivative with respect to time of equation 1, the energy rate is expressed as:

$$\frac{dE}{dt} = \Delta P = mV \frac{dV}{dt} + mg \frac{dh}{dt}. \quad (2)$$

2.1 REQUIRED POWER

The rotor power required in forward flight is given by the sum of parasite power, induced power, rotor blade profile power, compressibility power, stall power and climb power [14].

$$\begin{aligned} P_{req,rotor} = & 0.5f\rho V^3 + TV_i \\ & + 0.125\delta bcR(1 + 4.6\mu^2)\rho(\Omega R)^3 \\ & + \rho\Omega^3\pi R^5\Delta M^3[0.0033 - \Delta M(0.022 - 0.11\Delta M)] \\ & + \kappa^{1.5}(t_c - t_{div})^{1.5} + mgV_h, \end{aligned} \quad (3)$$

where f is equivalent flat-plate drag area, ρ is air density, T is rotor thrust, V_i is induced velocity, δ is coefficient of blade drag, b is number of blades, c is blade chord, R is rotor radius, μ is advance ratio, ΔM is the amount by which advancing blade tip Mach number exceeds drag divergent Mach number, κ is a constant coefficient, t_c is thrust coefficient, t_{div} is

*Corresponding author e-mail: luozp@buaa.edu.cn

thrust coefficient at which stall power occurs and V_h is vertical velocity. The total power required is obtained by rotor power and overall efficiency factor (η) and

$$HP_{req,total} = \eta HP_{req,rotor} \tag{4}$$

2.2 PERFORMANCE

Changes in horizontal velocity (V_l) for energy rate (ΔP_1) is determined form the following:

$$\frac{dV_l}{dt} = \frac{\eta_1 \Delta P_1}{mV_l} \tag{5}$$

where $\eta_1 = \begin{cases} 1 & , \text{ when } \Delta P_1 \geq 0 \\ 0.8 & , \text{ when } \Delta P_1 < 0 \end{cases}$

The relationship between normal load factor (n_n) and the turn rate ($\dot{\theta}$) is given by:

$$\dot{\theta} = \frac{g n_n}{V_l} \tag{6}$$

Energy rate (ΔP_2) influences vertical velocity as follows:

$$V_h = \frac{\eta_2 \Delta P_2}{m \frac{dV_h}{dt} + mg} \tag{7}$$

where $\eta_2 = \begin{cases} 1 & , \text{ when } \Delta P_2 \geq 0 \\ 0.8 & , \text{ when } \Delta P_2 < 0 \end{cases}$

2.3 ROTOR THRUST LIMITS

The maximum thrust of main rotor is restricted by the available power and maximum thrust coefficient.

$$t_c = \frac{T}{0.5 \rho b c \Omega^2 R^5} \leq t_{c,max} \tag{8}$$

$$\Delta P_1 + \Delta P_2 + (P_{req,total} - \eta mg V_h) \leq P_{ava} \tag{9}$$

where $t_{c,max}$ and P_{ava} is maximum thrust coefficient and available power.

2.4 KINEMATIC EQUATIONS

The kinematic equations can be written as follows:

$$\begin{cases} \dot{x} = V_l \cos \theta \\ \dot{y} = V_l \sin \theta \\ \dot{z} = V_h \end{cases} \tag{10}$$

3 Calculation algorithm

Figure 1 presents the algorithm of calculating the maneuverability and flight trajectory. The required data for this algorithm are properties of helicopters, engine output power and control laws.

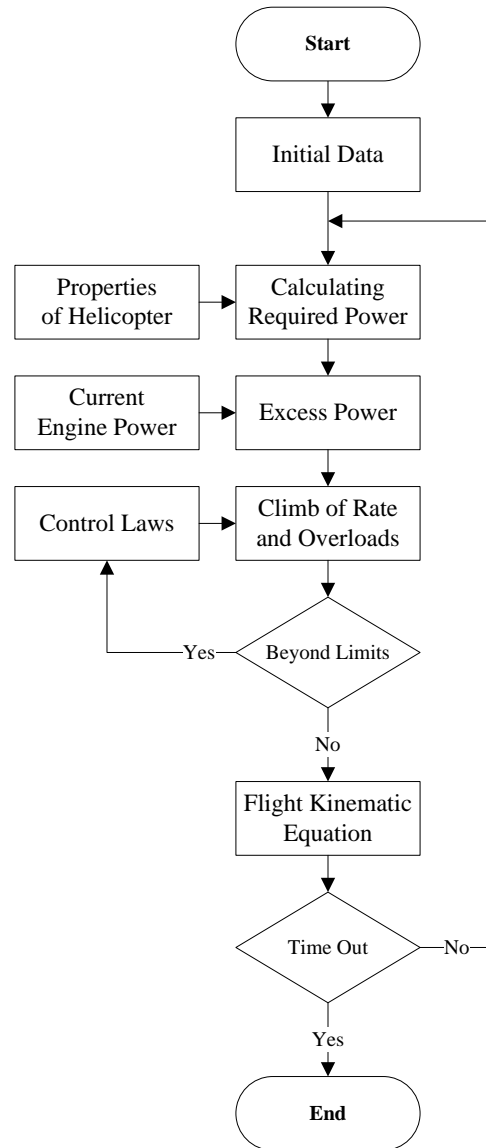


FIGURE 1 Calculation algorithm for maneuverability

4 Examples and results

A program is completed in MATLAB to calculate AH-1G helicopter maneuverability. The properties of AH-1G helicopter [15], which are used in program are presented in Table 1. Three kinds of maneuvers are analysed as follows.

TABLE 1 Properties of AH-1G helicopter

Parameter (unit)	Value	Parameter (unit)	Value
m (kg)	3400	Ω (s^{-1})	34
R (m)	6.71	f (m^2)	1.82
b (-)	2	κ (-)	736
c (m)	0.69	δ (-)	0.0075

4.1 LEVEL ACCELERATION

Level acceleration is a very necessary and integral component of the maneuver capability of a helicopter. It is very important to predict acceleration in danger situations. In the present example, the engine is at the maximum power output condition and the initial velocity is 25.7 m/s. Figure 2 shows the computed trajectory of the helicopter in 10 seconds. Figure 3 and Figure 4 demonstrate the time histories of the helicopter speed and power for the acceleration maneuver. It can be seen from the figures that the excess power is maximum at the velocity about 32 m/s.

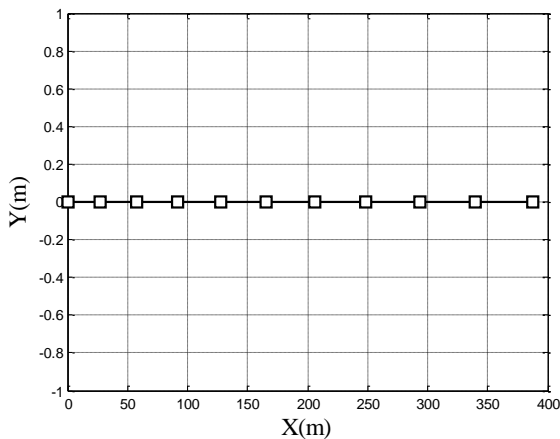


FIGURE 2 Acceleration trajectory

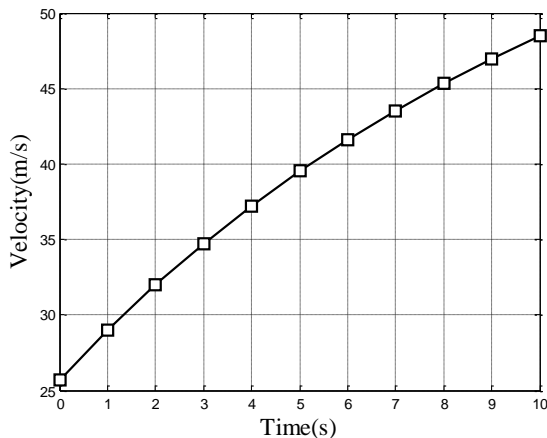


FIGURE 3 Time history of velocity for acceleration maneuver

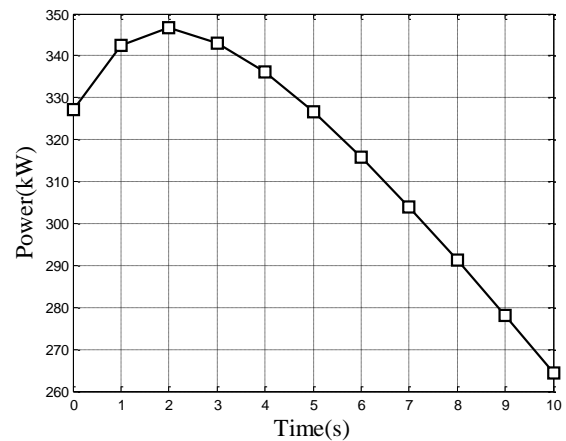


FIGURE 4 Time history of excess power for acceleration maneuver

4.2 DECELERATING TURN

The pilots often choose decelerate at constant altitude in order to supply more power for a 180 degree turn. Figure 5 presents an example of a decelerating turn at 1.72-g overload.

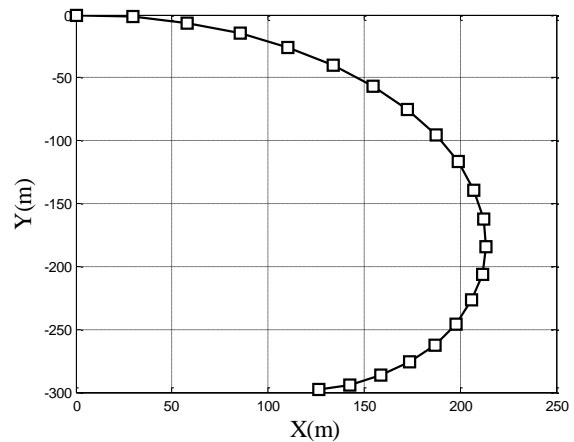


FIGURE 5 Turning trajectory

The time histories of the helicopter speed and power in this maneuver are given in Figure 6 and Figure 7. In a decelerating turn, the helicopter must maintain engine idle speed [16]. Therefore, the value of excess power is always below zero, shown in Figure 7. The maximum excess power is at the velocity about 37 m/s.

4.3 TURNING CLIMB FOLLOWED BY ACCELERATING CLIMB

A complex maneuver is presented in which a turning climb is following by an accelerating climb. The rate of climb is 5 m/s in whole maneuver. The helicopter maintains engine idle speed in climbing turn and the engine work at the maximum power output condition in accelerating climb.

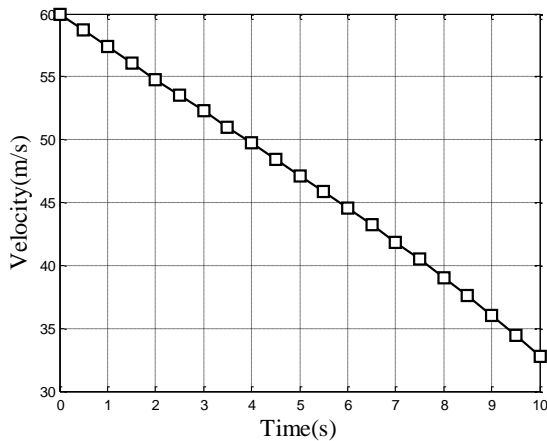


FIGURE 6 Time history of velocity for turning maneuver

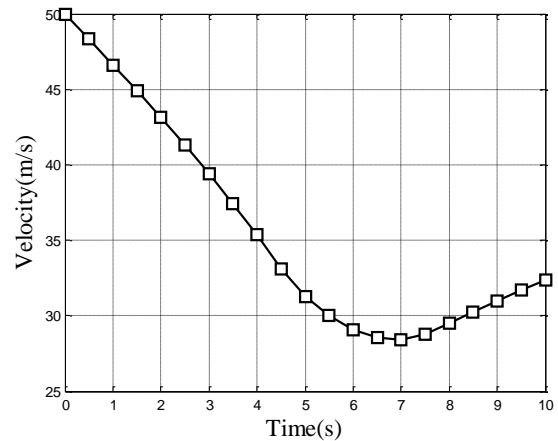


FIGURE 9 Time history of velocity for 3-D maneuver

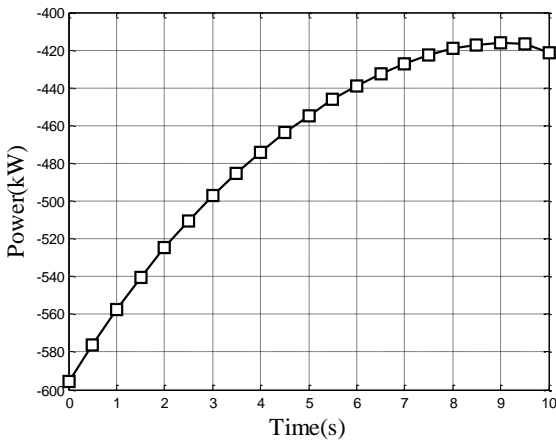


FIGURE 7 Time history of power for turning maneuver

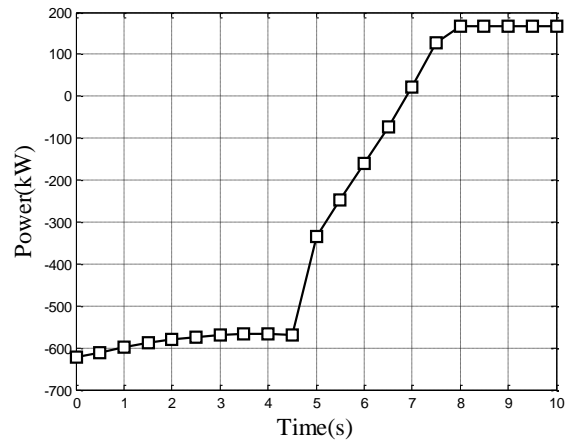


FIGURE 10 Time history of power for 3-D maneuver

There is a transitional period between these two conditions. Figure 8 shows the flight trajectory of this maneuver. Figure 9 demonstrates the time history of the helicopter speed. It can be seen from figure 10 that there is a transitional period about 3 seconds.

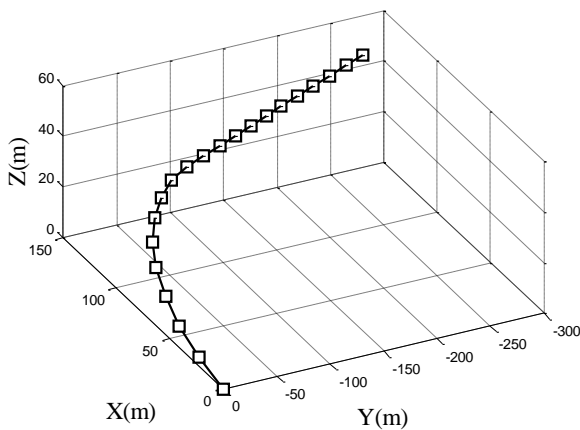


FIGURE 8 3-D flight trajectory

5 Conclusions

This paper proposed a fast and accurate approach to evaluate helicopter maneuverability in forward flight. A three dimensional complex maneuver of AH-1G helicopter is presented, and some important characteristics including maneuver time, trajectory, excess power and velocity are obtained. The results indicate that this method is feasible and effective for helicopters during forward flight. In order to get a fast calculation algorithm, some empirical studies which are only appropriate for forward flight situation are employed in mathematical model. Future research would focus on the rapid method for evaluating maneuverability of the helicopter in hover or vortex ring state.



Acknowledgments

This study has been supported by the State Key Basic Research Program of China (2011CB707002).

References

- [1] Meng S, Xiang J, Luo Z 2013 Navigation of micro aerial vehicle in unknown environments *25th Chinese Control and Decision Conference: Guiyang* 322-7
- [2] Li J, Gao Z 2009 A simplified algorithm of rotor power and pull to determine helicopter performance *Journal of Air Force Engineering University (Natural Science Edition)* 10(4) 9-12 (in Chinese)
- [3] Zhang Y 2003 Research of computing method for helicopter required power *Helicopter Technique* 22(1) 1-5 (in Chinese)
- [4] The General Editorial of Aircraft Design Manual 2005 *The 19th volume of aircraft design manual* Aviation Industry Press: Beijing (in Chinese)
- [5] Wells C D, Wood T L 1973 *Journal of the American Helicopter Society* 18(1) 10-22
- [6] Kiwan A R 1994 Helicopter performance evaluation (HELPE) computer model *AD Report AD-A284 319/1*
- [7] Xu X, Wang W 1994 Study on Z-9 helicopter turning maneuver characteristics by energy method *Journal of Nanjing University of Aeronautics and Astronautics* 26(1) 121-6 (in Chinese)
- [8] Mikhailov S A, Onushkin A Y 2007 Power balance method in calculation of helicopter maneuverability taking into account specific operational conditions. *Russian Aeronautics* 50(2) 121-8
- [9] Mikhailov S A, Onushkin Y P, Safonov A A, Kochish S I 2009 Numerical simulation of helicopter maneuverability in aerobatics research *Russian Aeronautics* 52(2) 176-83
- [10] Mikhailov S A, Onushkin A Y, Onushkin Y P, Safonov A A, Kochish S I 2009 Investigation of helicopter maneuverability by the power balance method *Russian Aeronautics* 52(3) 296-301
- [11] Cao Y, Zhang G, Su Y 2004 Mathematical modeling of helicopter aerobatic maneuvers *Aircraft Engineering and Aerospace Technology* 76(2) 170-78
- [12] Hu H, Zhang X 2012 Simulation method of helicopter maneuver flight based on mathematical description *Fire Control and Command Control* 37 99-101 (in Chinese)
- [13] Hu H, Wu S, Shao H 2011 Mathematical description and simulation of helicopter maneuver flight *System Simulation Technology* 7(4) 305-10 (in Chinese)
- [14] Johnson W 1994 *Helicopter theory* Dover Publications: New York
- [15] Wood T L, Livingston C L 1971 An energy method for prediction of helicopter maneuverability *AD Report, ADA021266*
- [16] Gao Z, Sun C 1997 Probing of helicopter maneuverability *Journal of Nanjing University of Aeronautics and Astronautics* 29(6) 666-73 (in Chinese)

Zhuang Nanjian, Xiang Jinwu, Luo Zhangping, Ren Yiru

Authors	
	<p>Nanjian Zhuang, born in June, 1987, Hunan, China</p> <p>Current position, grades: Ph.D. student at School of Aeronautic Science and Engineering, Beihang University, China.</p> <p>University studies: Bachelor degree, Beihang University, 2008</p> <p>Scientific interest: helicopter flight dynamic, structural dynamics.</p> <p>Publications: 2 papers</p>
	<p>Jinwu Xiang, born in February, 1964, Hunan, China</p> <p>Current position, grades: Professor at School of Aeronautic Science and Engineering, Beihang University, China.</p> <p>University studies: Bachelor degree (Nanjing University of Aeronautics and Astronautics, 1980), Master degree (Northwestern Polytechnical University, 1990), PhD (Nanjing University of Aeronautics and Astronautics, 1993).</p> <p>Scientific interest: aircraft design, aeroelasticity and structural dynamics, etc.</p> <p>Publications: 161 papers, 12 patents</p>
	<p>Zhangping Luo, born in December, 1968, Guizhou, China</p> <p>Current position, grades: associate professor at School of Aeronautic Science and Engineering, Beihang University, China.</p> <p>University studies: Bachelor degree (Beihang University, 1991), Ph. D. (Beihang University, 2004).</p> <p>Scientific interest: structural dynamics, impact dynamics and crashworthiness, aircraft design.</p> <p>Publications: 16 papers</p>
	<p>Yiru Ren, born in January, 1983, Hunan, China</p> <p>Current position, grades: post-doctor at School of Aeronautic Science and Engineering, Beihang University, China.</p> <p>University studies: Bachelor degree (Harbin Engineering University, 2006), Ph. D. (Beihang University, 2011).</p> <p>Scientific interest: structural dynamics, impact dynamics and crashworthiness, aircraft design.</p> <p>Publications: 10 papers</p>

Residual life prediction under condition monitoring

Shu Jie Liu^{1*}, Ya Wei Hu¹, Chao Li¹, Hong Chao Zhang^{1, 2}

¹School of Mechanical Engineering, Dalian University of Technology, Dalian, China

²Dept. of Industrial Engineering, Texas Tech University, Texas, America

Received 1 March 2014, www.tsi.lv

Abstract

Reliability assessment and remaining life prediction in the working processes of mechanical products, getting more attention of researchers, can reduce accidents and losses and help improve the preventive maintenance decision-making. This article presents two failure models, linear and exponential, to predict residual life distribution based on the degradation information of mechanical products. Parameters of the models can be estimated using maximum likelihood method. After the real-time monitoring information is acquired, residual life distribution should be updated constantly in order to improve accuracy of the prediction. Experiments were carried out on a double row cylindrical roller bearing to get the vibration information. It proved the validity of the aforementioned method and was applied to compare the two degradation models.

Keywords: reliability, residual life distribution, degradation model, double row cylindrical roller bearing

Introduction

In recent years, the reliability estimation study of mechanical parts or equipments have gained meaningful results at home and abroad. The feature information, which characterizes equipment's state, can be used to develop mathematical models. B E Cukor studied the combustion engines' wear rules of main parts under dynamic loading conditions and established mathematical models through regression of a large set of experimental data. F Y Wang et al carried out state detection on fans. Fitting process was conducted on the test data using least square method and the feasibility analyses of the limited operational life provided the expected effects [1]. A M Huang et al established a linear regression model of aviation hydraulic pump performance degradation with time using accelerated degradation test data and predicted the residual life [2]. The trend forecast based on Grey Model (GM) (1, 1) of grey system theory was introduced in [3, 4] and this article reported that the method was

reliable. R Y Li et al predicted the failure rate of aircrafts of some airlines based on Auto Regressive Moving Average (ARMA) model and illustrated that ARMA model was suitable for forecasting failure rate [5]. Lu and Meeker adopted ARMA model to predict the failure time, and estimated the distribution of the failure time [6]. On the basis of this work, N Gebrael et al developed an updated residual life distribution from the degradation signal in real time [7]. In general, the residual life prediction process based on the mathematical model can be shown as in Figure 1.

This paper establishes two models: the linear and the exponential. We assume that the distribution of the failure time is a Bernstein distribution. The distribution parameters are estimated using the maximum likelihood method and the mathematical models are gained to account for the remaining useful life (RUL) distribution. The residual life distribution can be updated by real-time data degradation continually.

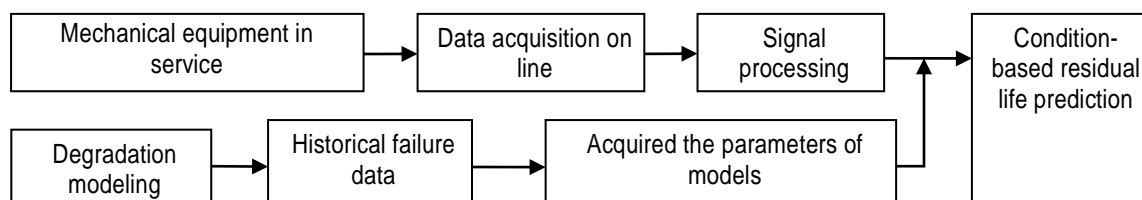


FIGURE 1 The process of residual life prediction

The models and the prediction algorithm

Establishing mathematical models by degradation signals, which can characterize the products' mechanical

properties, to forecast RUL has been widely focused. In the paper, the level of degradation signals at times t_i is defined as $S(t_i)$, $i = 1, 2, \dots$, and the mathematical model

*Corresponding author e-mail: liushujie@dlut.edu.cn

can be expressed as $S(t_i) = f(t_i; \phi, \beta) + \varepsilon(t_i)$, where $\varepsilon(t_i)$ is the error term assumed to be i.i.d. normal random variables with mean 0, and variance σ^2 . The deterministic parameter ϕ is a constant value in the whole degradation process. The parameter β is an assumed stochastic coefficient following a prior distribution. This paper describes two regression models and calculates the residual life distribution for each of them.

1.1 LINEAR DEGRADATION MODEL

The degradation model is given by the following equation

$$S(t_i) = \phi + \beta t_i + \varepsilon(t_i), \tag{1}$$

$$p(S_1, \dots, S_k / \beta) = \frac{1}{\prod_{i=1}^k \sqrt{2\pi\sigma^2}} \times \exp\left(-\sum_{i=1}^k \frac{(S_i - \beta t_i - \phi)^2}{2\sigma^2}\right). \tag{3}$$

Thus, the Eq. (3) is expressed as

$$p(\beta / S_1, \dots, S_k) \propto \exp\left(-\sum_{i=1}^k \frac{(S_i - \beta t_i - \phi)^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{(\beta - \mu_\beta)^2}{2\sigma_\beta^2}\right) \propto \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left(-\frac{(\beta - \mu_\beta)^2}{2\sigma_\beta^2}\right).$$

The parameters of the updated distribution of β are given by:

$$\mu_\beta = \frac{\sigma_\beta^2 \sum_{i=1}^k (S_i - \phi)t_i + \sigma^2 \mu_\beta}{\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2}, \text{ and } \sigma_\beta^2 = \frac{\sigma^2 \sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2}. \tag{4}$$

At some time level t in the future, the level of the degradation signal can be expressed by the random variables $S(t+t_k)$. The mean and the variance can be given by:

$$\mu(t+t_k) = \phi + \mu_\beta t; \sigma^2(t+t_k) = \sigma_\beta^2 t^2 + \sigma^2. \tag{5}$$

$$F(T_R \leq t | S_1, \dots, S_k) = P(S(t+t_k) \geq D | S_1, \dots, S_k) = \Phi\left(\frac{\mu(t+t_k) - D}{\sigma(t+t_k)}\right), \tag{6}$$

where $\Phi(\bullet)$ is the CDF of a standardized normal random variable. Factually, the value of a unit's residual life can never be negative, so the negative values of the remaining life should be precluded. Figure 2 is the

where the deterministic parameter $\phi = S(0)$. Assuming that $\pi(\beta)$ is the prior distribution of β , which is considered as i.i.d. $N(\mu_\beta, \sigma_\beta^2)$. We define $S_i = S(t_i)$, and assume that a unit degradation signal S_1, \dots, S_k has been observed up to a time level t_k . The conditional probability of the parameter β can be gained using Bayesian method and the updated distribution of β is

$$p(\beta / S_1, \dots, S_k) \propto p(S_1, \dots, S_k / \beta) \cdot \pi(\beta). \tag{2}$$

Because the error terms $\varepsilon(t_i)$ are i.i.d. $N(0, \sigma^2)$. So $(S(t_i) - \phi + \beta t_i)$ are following i.i.d. normal distribution with mean 0, and variance σ^2 . The joint distribution of the observed signal can be expressed as:

We define the predetermined failure threshold as D and the distribution of the residual life T_R can be determined from the conditional probability as follows:

probability density function of the residual life and the area of the shaded part should be eliminated. Thus, the CDF of the residual life is:

$$p(T_R \leq t | S_1, \dots, S_K) = \frac{\Phi(\frac{\mu(t+t_k)-D}{\sigma(t+t_k)}) - \Phi(\frac{\mu(t_k)-D}{\sigma(t_k)})}{1 - \Phi(\frac{\mu(t_k)-D}{\sigma(t_k)})} \tag{7}$$

2.2 EXPONENTIAL DEGRADATION MODEL

The exponential degradation model is widely used when the cumulative damage is significant [7, 8]. The form of the exponential model is defined as:

$$S(t_i) = \phi \exp(\beta t_i + \varepsilon(t_i) - \frac{\sigma^2}{2}), \tag{8}$$

where the parameter ϕ is a constant, and we assume that $\varepsilon(t_i)$ are i.i.d. normal random variables with mean 0, and variance σ^2 and the parameter β is a stochastic coefficient following a prior distribution, which is i.i.d. $N(\mu_\beta, \sigma_\beta^2)$, as in linear degradation model.

Now, we can transform the exponential model into a linear form as follows:

$$L_i = \ln S(t_i) = \phi' + \beta t_i + \varepsilon(t_i), \tag{9}$$

where $\phi' = \ln \phi - \frac{\sigma^2}{2}$, and ϕ' is also a constant.

To gain the parameters of the updated distribution of β is analogous to that of the linear modelling method introduced above and is a normal distribution. Similarly, the mean and the variance of the random variable $L(t+t_k)$ are

$$F(t) = p(T_L \leq t) = P(W(t) \geq D) = 1 - \Phi(\frac{D - (\mu_\lambda + \mu_\omega t)}{\sqrt{\sigma_\lambda^2 + \sigma_\omega^2 t}}), \tag{12}$$

where T_L is the unit's life, which is defined as a random variable. The unit is regarded to failure when $W(t)$ reaches the threshold Bernstein distribution is widely used in the residual life prediction. The failure time distribution in Eq. 14 can be expressed as a 2-parameter Bernstein distribution [9]:

$$L = \prod_{i=1}^n f(t_i) = \frac{c^n}{(2\pi a)^{n/2} \prod_{i=1}^n t_i^2} \exp(-\frac{1}{2a} \sum_{i=1}^n (1 - \frac{c}{t_i})^2). \tag{14}$$

After applying logarithm, it can be given by:

$$\ln L = n \ln c - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln a - 2 \sum_{i=1}^n \ln t_i - \frac{1}{2a} \sum_{i=1}^n (1 - \frac{c}{t_i})^2, \tag{15}$$

$$\mu(t+t_k) = \phi' + \mu_\beta t, \text{ and } \sigma^2(t+t_k) = \sigma_\beta^2 t^2 + \sigma^2. \tag{10}$$

3 The estimation of the prior distribution of the stochastic parameters

The prior distribution parameters should be estimated first when the residual life distribution of the mechanical products is updated. N. Gebraeel et al. demonstrate that there are almost no impacts on forecast accuracy of the models' parameters using aforementioned monitoring information and historical failure data [8]. This article introduces the computing method of the prior distribution parameters using historical failure data only. Degradation model can be expressed as:

$$W(t) = \lambda + \omega t, \tag{11}$$

where λ and ω are random variables and $W(t)$ is the level of degradation signals. We assume as $\lambda \sim N(\mu_\lambda, \sigma_\lambda)$ and $\omega \sim N(\mu_\omega, \sigma_\omega)$, thus, $W(t) \sim N(\mu_\lambda + \mu_\omega t, \sigma_\lambda^2 + \sigma_\omega^2 t)$.

The failure time distribution is expressed as:

$$f(t) = \frac{c}{\sqrt{2\pi a t^2}} \exp\{-\frac{1}{2a} (1 - \frac{c}{t})^2\}. \tag{13}$$

Now, estimate the stochastic parameters by maximum likelihood estimation method. The likelihood function is:

where $\{t_i\}$ is the set of failure times data.

The results are:

$$\hat{c} = \frac{n}{\sum_{i=1}^n 1/t_i}, \text{ and } a = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\log t_i} - \frac{1}{\hat{c}} \right)^2 \quad (16)$$

The parameters c and a are used to evaluate the mean μ_ω and variance σ_ω^2 of the parameter ω , and

$$\mu_\omega = \frac{D - \lambda}{c}, \text{ and } \sigma_\omega^2 = \mu_\omega^2 a \quad (17)$$

The estimated parameter ω is replaced by β , which is a random parameter of the established models (linear and exponential). Thus, the prior distribution of the degradation models is determined. The parameters of the updated distribution can be obtained according to the actual degradation process of the units and the residual life distributions based on the given conditions are obtained.

4 An example

The vibration data are taken from the Rexnord ZA-

2115 double row cylindrical roller bearings, using NI DAQCard-6062E data acquisition card, PCB 353B33 acceleration sensor, and the data acquisition software developed by LabVIEW. The sampling frequency is 20 kHz and the sampling length is 20480. It is required to calculate the vibration intensity through each of the 10min sampling data while the sampling interval is kept 10min. In order to verify the performance of the model, the model is established and analysed through one of the 180 vibration intensity data, which are shown in Figure 2. The vibration data are taken from the Rexnord ZA-2115 double row cylindrical roller bearings, using NI DAQCard-6062E data acquisition card, PCB 353B33 acceleration sensor, and the data acquisition software developed by LabVIEW. The sampling frequency is 20 kHz and the sampling length is 20480. It is required to calculate the vibration intensity through each of the 10min sampling data while the sampling interval is kept 10min. In order to verify the performance of the model, the model is established and analysed through one of the 180 vibration intensity data, which are shown in Figure 2.

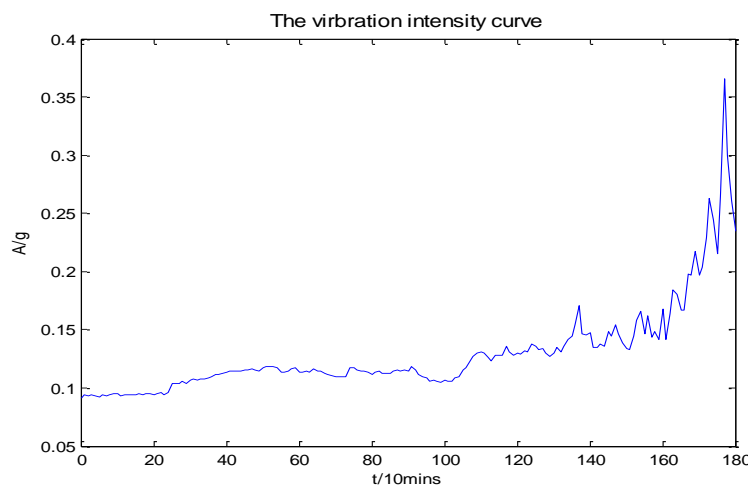


FIGURE 2 The vibration intensity curve

According to the modelling method and the calculation method described above, the prior distribution parameters of the two degradation models

(linear and exponential) are assessed using historical data of failure, which are shown in the Table 1.

TABLE 1 The estimated parameters of the prior distribution

	The value of $\phi (\phi')$	Distribution parameter of β	
		μ_β	σ_β^2
Linear Degradation model	0.0916	5.8661×10^{-4}	1.0978×10^{-6}
Exponential Degradation model	-2.4206	0.0044	1.9529×10^{-8}

The experimental results of the rolling bearing vibration intensity data are used to update the residual life distribution. Figure 3 illustrates evolution of the updated residual life distributions. Meanwhile, the

prediction errors of the linear model and the exponential model are compared and the results are shown in Figure 4. In the prediction, the median of the distribution of the residual life is regarded as estimated remaining life [8].

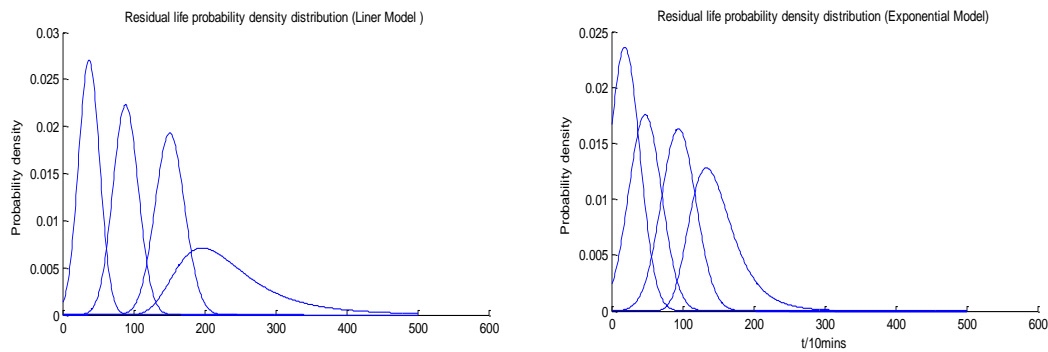


FIGURE 3 Evolution of residual life distributions

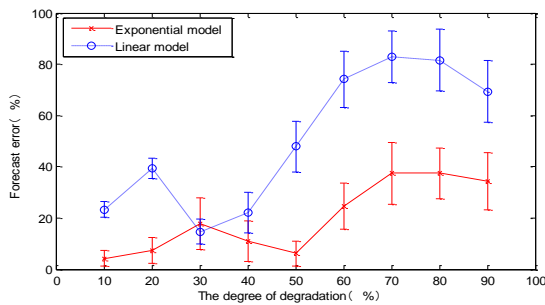


FIGURE 4 The errors of linear model and exponential model

From the experiment results, three observations are noteworthy:

- 1) The probability density distributions of the residual life are more and more concentrated with times. In the beginning, the probability density distributions are dispersive in both linear model and exponential models.
- 2) The prediction errors of the linear model are significantly larger than those of the exponential model, which we can be quantified from Figure 5. Because the rolling bearings failures are mostly caused by fatigue and the fatigue is a cumulative process. Thus, the exponential model is more suitable for the degradation process of the rolling bearings.
- 3) With increasing degree of degradation, the accuracy of the forecast seriously declines. When the bearing is close to failure, the intensity of vibrations is sharply increased, which is

References

[1] Wang F Y, Wang H Y, Su S D, Wang Y Z 1992 Estimate the limit life of blower with a simple diagnostic method *Equipment Management and Maintenance* 12 18-21 (in Chinese)
 [2] Huang A M, Guo Y E, Yu J F 2011 Research on residual life prediction technique of hydraulic pump based on accelerated degradation data *Machinery Design and Manufacture* (1) 154-5 (in Chinese)
 [3] Ma Z F, Li X F 2000 The Forecast for the Wear Trend of the Diesel Engine *Lubrication Engineering* (1) 54-6 (in Chinese)
 [4] Yin S Y, Chen D L 2005 Wear Trend Prediction for Aero-engine

observable in Figure 2. This implies that the exponential model is not suitable anymore for this stage.

5 Conclusions

This paper discusses a residual life prediction method of a unit or a system through adding online performance measurements continually. Two degradation models (linear and exponential) are established in the paper and the parameters of the two models are estimated by means of a widely used Bernstein distribution. The distributions of the residual life are updated with change of the units' state. The vibration intensity information extracted from the operation of double row cylindrical roller bearings validates the feasibility the method. Moreover, from comparison of the prediction errors, it can be concluded that the exponential model is more suitable than the linear model. Research on reliability of mechanical components or equipment can predict the future performance, which is helpful for preventive maintenance decision-making and improving quality of the system security assessment.

Acknowledgements

This work is jointly supported by the 973 Basic Research and Development Plan of China (No. 211CB013402) and the Special Fundamental Research Funds for Central Universities of China(No. DUT11RC(3) 60).

with GM(1,1) Model Method *Lubrication Engineering* 9(5) 96-7 (in Chinese)
 [5] Li R Y, Kang R 2008 Research on failure rate forecasting method based on ARMA model *Systems Engineering and Electronics* 30(8) 1588-91(in Chinese)
 [6] Lu H T, William J K, Lu S 2001 *Transactions Reliability* 50(4) 353-7
 [7] Gebraeel N, Lawley M, Li R, Ryan J K 2005 *IEEE Trans* 37(6) 543-57
 [8] Gebraeel N, Elwany A, Pan J 2009 *IEEE Transactions on Reliability* 58(1) 106-17

Authors	
	<p>Shujie Liu, born on August 25, 1977, Heilongjiang, China</p> <p>Current position, grades: Doctor of Engineering, Lecturer in Dalian University of Technology University studies: Mechanical Engineering in Jilin University Scientific interest: Surface profile measurement, Reliability and Life prediction Publications: 12 Papers Experience: 1999, graduate from school of mechanical engineering, Jilin University; 2007, obtain doctor degree in the University of Tokyo, Japan. Then, work successively as a postdoctoral researcher in Institute for Materials Research of Kyoto University and Tohoku University in Japan. 9, 2011. Join the faculty of mechanical engineering in Dalian University of Technology. Once involved in a number of advanced researches in the field of precision measurement, and received many awards.</p>
	<p>Yawei Hu, born on October 3, 1989, Shandong, China</p> <p>Current position, grades: Master of Mechanical Engineering, Dalian University of Technology University studies: School of Mechanical Engineering in Qingdao Agricultural University Scientific interest: Remaining useful life prediction, Condition based maintenance, Prognostic and health management and remanufacturing in Mechanical Engineering</p>
	<p>Chao Li, born on November 30, 1988, Shandong, China</p> <p>Current position: Master of Mechanical Engineering, Dalian University of Technology University studies: School of Mechanical Engineering in Dalian Jiaotong University Scientific interest: Remaining useful life prediction, Condition based maintenance in Mechanical Engineering</p>
	<p>Hongchao Zhang, born on August 2, 1953, Hebei, China</p> <p>Current position, grades: Doctor of Engineering, Dalian University of Technology as Thousands of people plan distinguished professor, professor of the Department of Industrial Engineering, Texas Tech University University studies: Mechanical Engineering in Technical University of Denmark Scientific interest: Green manufacturing, Remanufacturing and LCA Publications: 2 Books, 150 Papers Experience: 1989, obtain doctor degree from Technical University of Denmark. Then, work in Texas tech university. Meanwhile, participate in a number of national research activities. From 6, 2010, take office in Dalian University of Technology as Thousands of people plan distinguished professor. Currently, hold the post, lead scientist, in a national 973 project.</p>

Effects of global warming for building energy demand in China

Qing Ma^{1*}, Hua Yang², Chaogang Zhang³, Zhaohui Peng⁴

¹*School of Control Science and Engineering, Shandong University, Shandong, 250061, China*

²*Jining Tourism Administration, Shandong, 272000, China*

³*Shandong Yong'an real estate development company, Jinan, China*

⁴*Jinan HaoYuan System Engineering Co. Ltd., 250014, Jinan, China*

Received 1 March 2014, www.tsi.lv

Abstract

The impact of global warming on building energy demand in China was investigated by means of whole building energy analysis model and hourly weather data. Four standard multi-story office-building models, representative of four typical climate locations were constructed. For the time period 2050–2100, the climatic temperature scenario models for four typical cities was used that foresees a 2.7–4.2 °C rise in mean annual air temperature relative to the period 1961–1990 normal temperature and is thereby roughly in line with the climate change predictions made by the Intergovernmental Panel on Climate Change (IPCC). The simulation results show that the annual cooling energy demand for office buildings with internal heat gains of 20–30 W/m² will increase by 26–58% while the heating energy demand will fall by 17–52% for the period 2050–2100. This study has also shown that the typical meteorological year (TMY) currently in use by building designers and HVAC engineers in China will lead increasingly to an overestimation of heating energy demand. Similarly, the use of TMY to compute cooling power and cooling energy consumption is likely to result in a progressive underestimation of the future demand. The future building energy demand is set to become a crucial design issue.

Keywords: Global warming, Heating energy, Cooling energy, TMY, Temperature scenarios, Typical office building model

1 Introduction

The relationship between energy and climate is one of the hot issues concerned by all nations in the world. On one hand, the exploitation and utilization of energy affects climate. On the other hand, climate change also affects the energy uses. The fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC) predicts the most likely increment in surface temperature of 1.8–4.0 °C by the end of the 21 century, and growth scope of 0.3–6.4 °C [1]. Moreover, a wide band of uncertainty exists regarding the amount of warming. Recent research results from climate research institution in China suggest that averaged warming in China by the multi-model ensemble is in surface temperature of 1.9–5.5 °C by the end of the 21 century, a growth scope of 0.7–9.2 °C [1, 2]. Climatic parameters represent important boundary conditions for building design and the transient behaviour of the building envelope through its service life. Energy demand in buildings depends significantly on external boundary conditions, particularly on ambient temperature.

The impact of global warming on the energy consumption of a country for space heating and cooling depends on the current and future regional climate, the required thermal comfort inside buildings and technical building features such as thermal insulation quality. In

previous studies for the USA [2–4] and, more recently, for Greece [5] and Switzerland [6], climate change was found to have significant implications on energy consumption in buildings. Several methods have been proposed to estimate building energy demand from monthly temperature data. Thom [7] related US-HDD to the monthly average temperature and standard deviation of monthly temperature from its long-term average. His equation was later modified by Ref. [8]. Other methods first construct hourly weather data from monthly temperatures and then calculate degree-days from standard degree-day equations [9–12] use the Swiss standard for HDD and CDD to calculate the Swiss building energy demand. Unfortunately, all the above studies employed definitions of degree-days. Degree-day methods are simple procedures, only efficient procedures for constant heat gain in buildings, but for the most buildings, especially for commercial buildings, these estimations aren't accurate due to the fact that the internal temperature, thermal gains and building properties aren't relatively constant [14]. In all above studies, the building models were the old structures. In this paper, four standard office building models entirely meeting the requirement [15] are constructed. To our knowledge, no corresponding study has, so far, been attempted to adopt whole building energy analysis method for standard office buildings in China.

*Corresponding author e-mail: maging@sdu.edu.cn

2 Methodology

2.1 CLIMATE

The meteorological parameters in the paper are based on measured data (1971–2003) at 270 stations. Improvements were made in meteorological data. The climatic conditions looked at here are for the assessment of the performance of the building and associated HVAC systems. EnergyPlus, a computer program is used to measure the performance over a complete year at regular intervals up to one per hour. According to the requirement of criterion, the statistics year for outdoor air parameters is close to 30 years whereas building services engineers use a single year for simulation modelling. It is therefore important that the year chosen is representative of the weather over a number of years. Such a year is commonly called a typical meteorological year (TMY). The method used to select a TMY differs between countries; however, the objective is the same; to construct a set of 12 months that is representative of the past years (say 30 years). This means that a TMY is unlikely to include extremes and therefore, while suitable for the prediction of energy consumption it is unsuitable for the purpose of assessing the performance of buildings under more onerous conditions. The latter requires a year that contains periods when temperatures are higher than average.

Weather data for four cities representative of four typical climatic conditions in China were chosen for this study. Hourly weather data during the period from 1971 to 2003 were analysed. Meteorological data show typical meteorological year for the four typical cities according to the last 30-year weather data (1971–2003).

2.2 TEMPERATURE SCENARIOS

The irreducible uncertainties associated with the future global socio-economic development [13] make any projections of future climate change inherently difficult. As a result, the aim of this study was not to determine the “most likely” trajectory for future building energy demand in China but rather to mark out the range of possible futures. For this purpose, Model for the Assessment of Greenhouse gas Induced Climate Change (MAGICC) that drives a spatial Global and Regional Climate SCENario GENerator (SCENGEN) were first compiled. MAGICC has been the primary model used by IPCC to produce projections of future global-mean temperature and sea level rise. The climate model in MAGICC is an upwelling-diffusion energy-balance model that produces global- and hemispheric-mean output. All scenarios were based on simulation results from so-called coupled Atmosphere–Ocean General Circulation Models (AO-GCM). AO-GCM is the most sophisticated tools currently available to project possible changes in global climate. In SCENGEN, the globe is discretized such that temperature, rainfall etc. are

computed on a global grid. The horizontal grid point distance is in the order of five degrees. To use MAGICC / SCENGEN program for the climate warming, two main steps were considered: (i) Running MAGICC, Emissions Scenarios and Model Parameters in MAGICC are edited and (ii) Running SCENGEN and setting up analysis, model, region and variable.

In the first step, running MAGICC, the following emissions scenarios will be used: P50 for the Reference case and WRE350 for the Policy case. P50 is the median of the SRES emissions scenarios. WRE350 is the same as P50 except for CO₂ emissions, which are modified to follow the WRE350 concentration profile.

In the second step, on ‘region’ window, the four typical cities can be selected from a range of ‘hard-wired’ regions. The latitude/longitude domain will be shown numerically on the right. After setting up other parameters, the result shows the four cities change (Table 5) in annual-mean temperature for the 30-year interval centred on 2080 (for the P50 emissions scenario, and ‘best guess’ climate model parameters in MAGICC) averaged over all 17 AO-GCM in the SCENGEN model data base. The changes correspond to a global-mean warming of 2.4 °C, and the patterns include the effects of aerosols according to the aerosol selection made in MAGICC.

TABLE 1 The four cities change in annual-mean temperature for the 30-year interval centred on 2080

City	Harbin	Beijing	Shanghai	Guangzhou
Longitude (deg)	126.77E	116.47E	121.45E	113.33E
Latitude(deg)	45.75N	39.8N	31.40N	23.17N
Elevation (m)	142.3	31.3	5.5	41.0
Change (deg C)	4.2	4.1	3.7	2.7

2.3 STANDARD OFFICE BUILDING MODEL

The built stock in China consists of some 5 hundreds million good-sized buildings. According to the standard, four 80 m long, 25 m wide and 35 m tall heavyweight multi-story building (Figure 1) was modelled for the study presented in this paper.

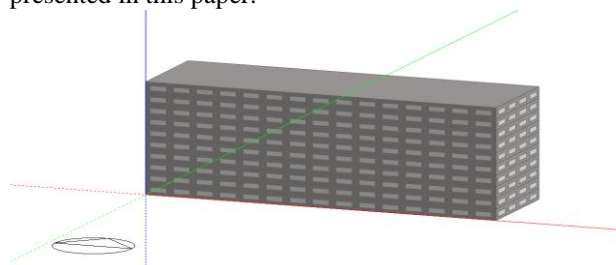


FIGURE 1 Model of the standard multi-story office building

The four standard building models meet the requirement of energy-saving building completely. The thermal insulation levels used to represent the four cities building code framework are given in Table 2.

The internal sensible heat gains considered are set out in Table 3. The per-capita area of use for the different type rooms is listed in Table 4.

TABLE 2 Thermal values of building components (shape factor<0.3, window wall area ratio=0.3) (W/(m² K)) [Design standard, 2005]

City	Uroof	Uouterwall	Uwindow
Harbin	0.35	0.45	2.8
Beijing	0.55	0.6	3.0
Shanghai	0.7	1.0	3.5
Guangzhou	0.9	1.5	4.7

TABLE 3 Internal heat gains for the different type rooms (W/m²) [Design standard, 2005]

Room sort	General	High	Meeting	Corridor	Other
Illumination power	11	18	11	5	11
Electrical equipment power	13	20	5	0	5

TABLE 4 The per-capita area of use for the different type rooms (m²/person) [Design standard, 2005]

Room sort	General	High	Meeting	Corridor	Other
per-capita area of use	4	8	2.5	50	20

The air conditioning systems were assumed to adopt the primary air and fan-coil systems that are used to design office buildings in China. The cold and heat source adopt the electric screw chiller and gas-fired boiler. The volume of primary air was designed for main space at 30m³ / (h·p).

2.4. SIMULATION MODEL AND ASSUMPTIONS

The EnergyPlus program is a collection of many program modules that work together to calculate the energy required for heating and cooling a building using a variety of systems and energy sources. It does this by simulating the building and associated energy systems when they are exposed to different environmental and operating conditions. The core of the simulation is a model of the building that is based on fundamental heat balance principles. It is based on the assumptions that the air in the thermal zone, by default, has a uniform temperature, the temperature of each surface is uniform, the long- and short-wave irradiation is uniform, the surface irradiation is diffusive and the heat conduction through the surfaces is one-dimensional. The formulae of the solution scheme starting with a heat balance on the zone are follows:

$$C_z \frac{dT_z}{dt} = \sum_{i=1}^{N_{zi}} \dot{Q}_i + \sum_{i=1}^{N_{surface}} h_i A_i (T_{si} - T_z) + \sum_{i=1}^{N_{zones}} \dot{m}_{inf} C_p (T_{zi} - T_z) + \dot{m}_{inf} C_p (T_{si} - T_z) + \dot{Q}_{sys}, \quad (1)$$

where N is the number of convective internal loads, \dot{Q}_i , $h_i A_i (T_{si} - T_z)$ is the convective heat transfer from zone surfaces at temperature T_{si} , while $\dot{m}_{inf} C_p (T_{zi} - T_z)$ is the heat transfer due to ventilation with the outside air, $\dot{m}_{inf} C_p (T_{si} - T_z)$ is the heat transfer due to inter zone air

mixing, and \dot{Q}_{sys} is the system output. The capacitance C_z takes into account the contribution of the zone air as well as that of thermal masses, which are assumed to be in equilibrium with the zone air.

If the air capacitance is neglected, the steady state system output must be:

$$-\dot{Q}_{sys} = \sum_{i=1}^{N_{zi}} \dot{Q}_i + \sum_{i=1}^{N_{surface}} h_i A_i (T_{si} - T_z) + \sum_{i=1}^{N_{zones}} \dot{m}_{inf} C_p (T_{zi} - T_z) + \dot{m}_{inf} C_p (T_{si} - T_z). \quad (2)$$

The multi-story office building was partitioned into three non-conditioned thermal zones, a zone representing the first floor, a zone representing the roof and a zone representing each single office. In order to cancel out the effects of system intermittency on the energy demand, a continuous operating schedule of the conditioning system was considered in all cases, assuming a heating set point of 18 °C and a cooling set point of 26 °C.

2.5 SIMULATION OF CURRENT AND FUTURE BUILDING ENERGY DEMAND

The simulation of current building energy demand was actualized according to the TMY and standard building

model established above, but simulation of future energy demand needs adding the temperature scenarios. As stated above, the limitations imposed by the annual-mean temperature scenarios forced us to base all our calculations for the season energy demand. However, the associated loss in accuracy appeared is tolerable. We were more interested in assessing long term changes and trends than precisely predicting individual monthly energy demand values. In this paper, a whole building energy analysis method was developed, comprising the following two steps.

In the first step, the building model was established in EnergyPlus. Through programming software, we can transform the XSL file into IDF file, thus the TMY

weather data can be used for weather data in EnergyPlus. It was executed for the simulation of current building energy demand to explore the difference of four cities energy demand. In the second step, the future energy demand was simulated. Based on the first step, the temperature scenarios must be added to forecast the change of the future energy demand.

3 Results

3.1 IMPACT ON HEATING AND COOLING ENERGY DEMAND

The energy demand for heating decreased significantly in Harbin, Beijing, Shanghai and Guangzhou in 2080 (Figure 3). Figure 3 includes result of simulation for heating in 2007 and 2080. In the four cities, the heating energy demand is very different in 2007. Harbin is most, while Guangzhou is least. For 2080, a 17% heating energy reduction was determined for Harbin (+4.2°C temperature rise), 28% drop for Beijing (+4.1°C temperature rise), 34% drop for Shanghai (+3.7°C temperature rise) and 52% drop for Guangzhou (+2.7°C temperature rise). The cooling energy demand in 2007 and 2080 is shown in Figure 2. It can be seen that the cooling energy demand rose at all locations, with the largest rates of increase occurring in Beijing and Harbin. The increase in cooling energy demand for climate scenario (+4.2°C temperature rise) is 45% for Harbin, 58% rise for Beijing (+4.1°C temperature rise), 34% rise for Shanghai (+3.7°C temperature rise) and 26% rise for Guangzhou (+2.7°C temperature rise).

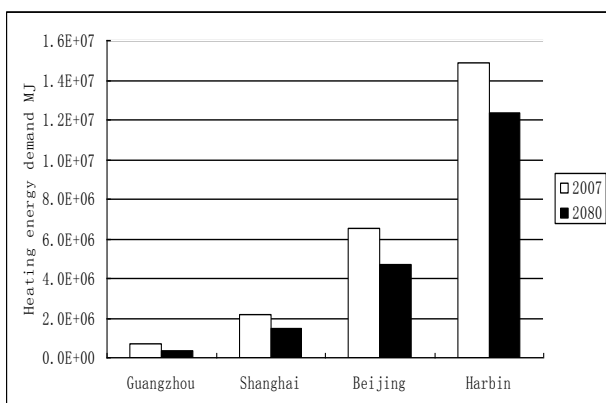


FIGURE 2 Annual heating energy demand for four typical cities in 2007 and 2080

3.2 IMPACT ON BUILDING DESIGN

Today, building designers in China use the TMY for the building design. The TMY was chosen from 30 years monthly meteorological data provided by the Weather

Bureau in China. A comparison between the heating and cooling energy demand in 2007 and 2080 is presented in Figure 2 and Figure 3. It can be seen that the difference of current and future TMY, which are based on measurements in the periods 1973–2003 and temperature scenarios, increasingly overestimated the heating energy demand and underestimated the cooling energy demand. This will affect building designer for future building design.

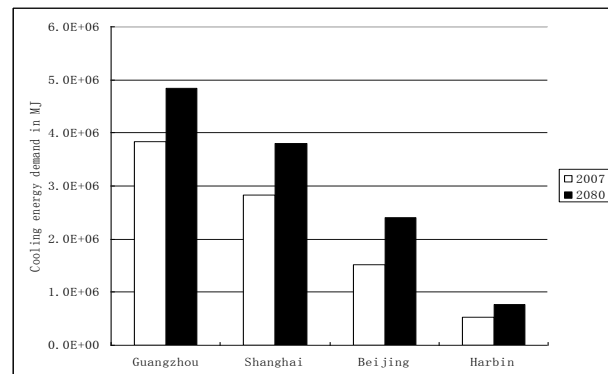


FIGURE 3 Annual cooling energy demand for our typical cities in 2007 and 2080

4 Conclusions

The impact of climate change on energy demand for heating and cooling was investigated in this study. Application of the whole building energy analysis method with four standard multi-story office building models, representative of four typical climate locations revealed a significant relative heating energy demand decrease and cooling energy demand increase, over the period 2007–2080. This relative decrease, between 17% and 52% depending on building location, was most pronounced at colder regions like Harbin. On the other hand, the cooling relative increase, between 26% and 58% depending on building location, was highest at the hottest site, Guangzhou. The four standard building models meet the requirement of energy-efficiency. The change in energy demand is more obvious if the thermal insulation level in building is worse.

Buildings in China have a long lifespan of about 50–100 years. This study has also shown that the TMY currently in use by building designers and HVAC engineers in China will lead increasingly to an overestimation of heating energy demand. Similarly, the use of TMY data to compute cooling power and cooling energy consumption is likely to result in a progressive underestimation of the future demand. It, therefore, seems obvious that continuous updating of weather data for building design is needed. The presented paper aims to initiate a discussion of this issue.

References

[1] Zhao Z C, Wang S W, Luo Y 2007 Assessments and projections of temperature rising since the establishment of IPCC *Advances in climate change* 3(3) 183-4 (in Chinese)

[2] Christenson M, Manz H, Gyalistras D 2004 Climate warming impact on degree-days and building energy demand in Switzerland *Energy Conversion and Management* 47 671-86

[3] Rosenthal D H, Gruenspecht H K, Moran E A 1995 Effects of global warming on energy use for space heating and cooling in the United States *The Energy Journal* 16(2) 77-96

[4] Belzer D B, Scott J M, Sands R D 1996 Climate change impacts on U.S. commercial building energy consumption: An analysis using sample survey data *Energy Sources* 18(2) 177-201

[5] Pretlove S E C, Oreszczyn T 1998 Climate change: impact on the environmental design of buildings *Proc CIBSE A Build Serv Eng Res Technol* 19 55-8

[6] Cartalis C, Synodinou A, Proedrou M, Tsangrassoulis A, Santamouris M 2001 Modifications in energy demand in urban areas as a result of climate changes: an assessment for the southeast Mediterranean region. *Energy Conv. Manag.* 42 1647-56

[7] Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S 2000 *IPCC special report on emissions scenarios* Cambridge: Cambridge University Press 319 pp

[8] Mearns L O, Hulme M, Carter T R, Leemans R, Lal M, Whetton P 2001 Climate scenario development *Climate change. The Scientific Basis* eds Houghton J T, Ding Y, Griggs D J, Noguer M, van der Linden P J X Dai et al Cambridge: Cambridge University Press 739 pp

[9] Thom H C S 1954 Normal degree-days below any base *Month Weather Rev.* 82(5) 111-5

[10] Erbs D G, Klein S A, Beckman, W A 1983 Estimation of degree-days and ambient temperature bin data from monthly-average temperatures *ASHRAE Journal* 25(6) 60-6





[11] Cannistraro G, Giaconia C, Pietrafesa M, Rizzo G 1995 Reduced weather data for building climatization and application to 29 European locations *Energy* 20(7) 637-46

[12] Gelezenis J J 1999 Estimation of hourly temperature data from their monthly average values: case study of Greece *Renewable Energy* 18(1) 49-60

[13] Badescu V, Zamfir E 1999 Degree-days, degree-hours and ambient temperature bin data from monthly-average temperatures (Romania) *Energy Conversion & Management* 40 885-900

[14] *EnergyPlus Engineering Reference* Washington, DC: US Department of energy <http://www.energyplus.gov>.

[15] *MAGICC/SCENGEN Program*, <http://www.cgd.ucar.edu>

Authors	
	<p>Qing Ma</p> <p>Current position, grades: Associate Professor at Shandong University, Jinan, Shandong, China.</p> <p>University studies: PhD degree on Control Theory and Control Engineering from Shandong University of China, 2012.</p> <p>Scientific interest: research interest: mainly focuses on computer network, data mining, the application of kinds of labels.</p>
	<p>Hua Yang</p> <p>Current position, grades: Shandong Yong'an real estate development company.</p> <p>University studies: Received bachelor degrees on electronic technology and computer application from Xi'an Electronic Technology University of China, 1994</p> <p>Scientific interest: environment engineering</p>
	<p>Chaogang Zhang</p> <p>Current position, grades: Jinan HAOYUAN system engineering Co Ltd</p> <p>Research interest: electrical technology, environment engineering</p>
	<p>Zhaohui Peng</p> <p>Current position, grades: Jining Tourism Administration</p> <p>University studies: bachelor degrees on Control Theory and Control Engineering from Shandong Industry University of China, 1994</p> <p>Research interest: environment engineering</p>

On the error rate analysis of dual-hop relaying over composite fading channels using mixture gamma distribution

Weijun Cheng*

School of Information Engineering, Minzu University of China No. 27 South Zhongguancun Street, Haidian District, Beijing 100081, China

Received 1 March 2014, www.tsi.lv

Abstract

In this paper, we present the end-to-end performance of a dual-hop amplify-and-forward (AF) relaying system over independent non-identical (non-i.i.d) composite Nakagami-lognormal (NL) fading channels by using mixture gamma (MG) distribution. Novel closed-form expressions for the probability density function (PDF) and the moment-generation function (MGF) of the end-to-end signal-to-noise ratio (SNR) are derived. Moreover, the average error rate and the diversity order are found based on the above new expressions, respectively. These expressions are more simple and accuracy than the previous ones obtained by using generalized-K (KG) distribution. Finally, numerical and simulation results are shown to verify the accuracy of the analytical results. These results show that it is more precise to approximate the composite NL distribution by using the MG distribution than using the KG distribution in the performance analysis of cooperative relaying systems.

Keywords: Dual-hop Relaying, Nakagami-lognormal, Mixture Gamma Distribution, Error Rate Analysis

1 Introduction

Multihop cooperative transmission has emerged as a promising technique for extending coverage, enhancing connectivity, and saving transmitter power in wireless communication networks. In the past decade, the performance analysis of the multihop cooperative system in term of outage probability and average bit/symbol error rate (BER/SER) has been widely studied over various multipath fading models, such as, Rayleigh [1], Nakagami-m [2], Rician [3], Weibull [4] and Generalized Gamma [5].

Recently, as an approximation fading model of composite multipath/shadowing fading channels, the generalized-K (KG) channel model has attracted considerable attention in the performance analysis of the multihop cooperative system. In [6], the performance of a dual-hop relaying system with fixed-gain relays was obtained. In [7], authors studied the error rate performance of multiple dual-hop relaying with maximum ratio combining. In [8], the performance of dual-hop relaying with best relay selection are evaluated. Authors in [9, 10] presented also their analysis of a dual-hop system.

Unfortunately, since their probability density functions (PDF) include modified Bessel functions, their cumulative distribution functions (CDF) and moment-generation functions (MGF) usually include some more complicated special functions, such as Meijer's G functions. To avoid mathematical difficulties, some further approximation has to be used. In [5] and [8], the Padé approximants (PA) method has been employed to

obtain the BER/SER. [6-9] adopted minimum SNR approximation model under amplify-and-forward (AF) strategy.

In [11, 12], the authors have developed an alternative approach by using the Mixture Gamma (MG) distribution to approximate the Nakagami-lognormal (NL) distribution. This distribution avoids the above-mentioned problems, and some exact and simple results obtained are possible by adjusting some parameters. To the best of our knowledge, there are few papers in performance analysis of cooperative relaying systems over independent non-identical (non-i.i.d) NL fading channels by using MG distribution.

In this paper, we examine the end-to-end performance of a dual-hop AF relaying system over non-i.i.d MG fading channels. Firstly, some novel closed-form expressions for the PDF and MGF of the end-to-end SNR are derived over MG fading channels, and some approximate results are obtained over KG fading channels for the purpose of comparison. Secondly, these new statistical results are used to evaluate the performance criteria of the dual-hop relay system. Finally, the numerical and simulation results are discussed and compared over MG and KG fading channels, and these results demonstrate the validity of the proposed analysis.

The remainder of this paper is organized as follows. In the next section, three models of composite fading channels are described, respectively. In section 3, a dual-hop relaying system is presented and the statistics of the end-to-end SNR are derived, such as the PDF and the MGF. Section 4 gives several important performance

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

criteria. Numerical and simulations results are presented in Section 5. Finally, we conclude our paper in section 6.

2 The Composite Fading Channel Model

2.1 NL FADING CHANNEL MODEL

For the NL channel, the composite NL distribution is a mixture of Nakagami and lognormal shadowing. Let X be the fading amplitude, which is a random variable. Assuming that the transmitted symbol energy is E_s and single sided power spectral density of the complex additive white Gaussian noise is N_0 , the instantaneous SNR per symbol is $\gamma = X^2 \rho$, where $\rho = E_s/N_0$ denotes the un-faded SNR.

The PDF of γ over NL channel, $f_{\gamma_{NL}}(\gamma)$, is given by [13]:

$$f_{\gamma_{NL}}(\gamma) = \int_0^\infty \frac{m^m \gamma^{m-1} \exp(-m\gamma/\rho y)}{\Gamma(m)(\rho y)^m} \frac{1}{\sqrt{2\pi\lambda y}} \exp\left[-\frac{(\ln y - \mu)^2}{2\lambda^2}\right] dy \quad (1)$$

where m is fading parameter in Nakagami fading, μ and λ are the mean and the standard deviation of lognormal shadowing, respectively, $\Gamma(\cdot)$ is the gamma function. Since a closed-form expression of (1) is not available in the literature, some approximations or simple forms of (1) have been given a great attention recently, such as, KG and MG distribution.

2.2 KG FADING CHANNEL MODEL

As an approximation of NL distribution, KG distribution is a mixture of Nakagami- m and Gamma distributions, where Gamma distribution approximates lognormal distribution. For KG fading, the PDF of γ , $f_{\gamma_{KG}}(\gamma)$, is given by [14]

$$f_{\gamma_{KG}}(\gamma) = 2\Xi^{(k+m)/2} \gamma^{(k+m)/2-1} K_{k-m}(2\sqrt{\Xi\gamma})/\Gamma(m)\Gamma(k), \quad (2)$$

where k is the distribution shaping parameter reflecting shadowing effect, $\Xi = km/\bar{\gamma}$, $\bar{\gamma} = k\Omega\rho$, Ω is the local mean power, $K_\alpha(\bullet)$ is the second kind modified Bessel function of order α defined in [15].

For integer values of m and arbitrary values of k , The CDF of γ over KG channel, $F_{\gamma_{KG}}(\gamma)$, can be expressed as in [6]:

$$F_{\gamma_{KG}}(\gamma) = 1 - \sum_{i=0}^{m-1} 2(\Xi\gamma)^{(k+i)/2} K_{k-i}(2\sqrt{\Xi\gamma})/\Gamma(k)i!. \quad (3)$$

Moreover, the MGF of γ over KG channel is given by [6]:

$$M_{\gamma_{KG}}(s) = G_{2,1}^{1,2}\left(s/\Xi \middle| 1-k, 1-m\right) / \Gamma(m)\Gamma(k), \quad (4)$$

where $G(\bullet)$ is the Meijer's G-function defined in [15].

2.3 MG FADING CHANNEL MODEL

By using Gaussian-Hermite quadrature sum, an alternative approximating approach of (1) is firstly given by [12], named as MG distribution. As in [12], the PDF, CDF and MGF of γ over MG channel can, respectively, be written as:

$$f_{\gamma_{MG}}(\gamma) = \sum_{i=1}^N C a_i \gamma^{m-1} \exp(-\frac{b_i}{\rho} \gamma) / 2\rho^m, \quad (5)$$

$$F_{\gamma_{MG}}(x) = 1 - \sum_{i=1}^N C a_i \Gamma(m, \frac{b_i}{\rho} x) / 2b_i^m, \quad (6)$$

$$M_{\gamma_{MG}}(s) = \sum_{i=1}^N C \Gamma(m) a_i / [2\rho^m (s + b_i/\rho)^m], \quad (7)$$

where $a_i = 2m^m w_i \exp[-m(\sqrt{2}\lambda t_i + \mu)] / \sqrt{\pi} \Gamma(m)$, $b_i = m \exp[-(\sqrt{2}\lambda t_i + \mu)]$, 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 [16, Table(25.10)] or can be calculated by a simple MATLAB program. $\Gamma(\bullet, \bullet)$ is the incomplete gamma function defined in [15].

For the corresponding relationships between parameters (μ and λ) in NL/MG model and parameters (k and m) in KG model, two approximating expressions can be derived as [17], i.e., $\lambda^2 = \Psi'(k)$, $\mu = \Psi(k) + \ln(\Omega)$, where $\Psi(\bullet)$ is the first derivative of psi function defined in [15].

3 The Dual-hop AF Relaying System Model

We consider a wireless dual-hop AF relaying system, which the source-destination link is unavailable. The source node (S) communicates with the destination node (D) over a relay node (R). It is assumed that nodes are synchronized and the channel state information is available at the receivers (R and/or D).

Let the instantaneous SNR of each hop link be denoted by γ_i , $i \in \{1, 2\}$. Similar as in [1] and [2], the instantaneous equivalent end-to-end SNR of the dual-hop link at the destination node can be expressed as:

$$\gamma_{SRD} = \gamma_1 \gamma_2 / (\gamma_1 + \gamma_2 + 1), \quad (8)$$

where $\gamma_i = X_i^2 \rho$, X_i is the fading amplitude of the i^{th} -hop link. The corresponding average SNR is defined as $\bar{\gamma}_i = \Omega_i \rho$. Due to capture the path-loss effect, $\Omega_i = E[X_i^2] = (d_0/d_i)^\epsilon$, d_i denotes the distance of the i^{th} -hop link, d_0 denotes the distance between S and D, and ϵ is the path-loss exponent.

In order to simplify the performance analysis of (8) over Nakagami-m, Weibull and KG fading, its upper bound is often adopted in many recent literature as [4, 6-9]:

$$\gamma_{SRD} < \gamma_b = \min(\gamma_1, \gamma_2). \tag{9}$$

This upper bound has been shown to be accurate enough at medium and high SNR values. For the purpose of comparison, we consider some statistics of (9) in this paper.

$$f_{\gamma_{b-MG}}(\gamma) = \sum_{i=1}^N \sum_{j=1}^N \frac{C^2 a_i a_j}{4 \rho^{m_i} b_j^{m_2}} \gamma^{m_1-1} \exp\left(-\frac{b_i}{\rho} \gamma\right) \Gamma(m_2, \frac{b_j}{\rho} \gamma) + \sum_{i=1}^N \sum_{j=1}^N \frac{C^2 a_i a_j}{4 b_i^{m_1} \rho^{m_2}} \gamma^{m_2-1} \exp\left(-\frac{b_j}{\rho} \gamma\right) \Gamma(m_1, \frac{b_i}{\rho} \gamma), \tag{11}$$

By using (11), the MGF of γ_b can be expressed with the aid of (6.621.3) in [15] as:

$$M_{\gamma_{b-MG}}(s) = \sum_{i=1}^N \sum_{j=1}^N \frac{C^2 a_i a_j \Gamma(m_1 + m_2)}{4 m_1 (b_i + b_j + \rho s)^{m_1 + m_2}} {}_2F_1(1m_1 + m_2; m_1 + 1; \frac{b_i + \rho s}{b_i + b_j + \rho s}) + \sum_{i=1}^N \sum_{j=1}^N \frac{C^2 a_i a_j \Gamma(m_1 + m_2)}{4 m_2 (b_i + b_j + \rho s)^{m_1 + m_2}} {}_2F_1(1m_1 + m_2; m_2 + 1; \frac{b_j + \rho s}{b_i + b_j + \rho s}), \tag{12}$$

where ${}_2F_1(a, b; c; z)$ is the hypergeometric function defined in [15].

3.2 STATISTICS OVER KG FADING MODEL

Similar as (11), and using (2) and (3), the PDF of γ_b over non-i.i.d KG fading channel can be found as:

$$f_{\gamma_{b-KG}}(\gamma) = \sum_{i=0}^{m_2-1} \frac{4 \Xi_1^{(k_1+m_1)/2} \Xi_2^{(k_2+i)/2}}{i! \Gamma(m_1) \Gamma(k_1) \Gamma(k_2)} \gamma^{\frac{k_1+m_1+k_2+i}{2}-1} K_{k_1-m_1}(2\sqrt{\Xi_1} \gamma) K_{k_2-i}(2\sqrt{\Xi_2} \gamma) + \sum_{i=0}^{m_1-1} \frac{4 \Xi_2^{(k_2+m_2)/2} \Xi_1^{(k_1+i)/2}}{i! \Gamma(m_2) \Gamma(k_2) \Gamma(k_1)} \gamma^{\frac{k_2+m_2+k_1+i}{2}-1} K_{k_2-m_2}(2\sqrt{\Xi_2} \gamma) K_{k_1-i}(2\sqrt{\Xi_1} \gamma). \tag{13}$$

where $k_i (i=1,2)$ is the distribution shaping parameter of the i -th-hop link, $\Xi_i = k_i m_i / \bar{\gamma}_i$.

In order to reduce the difficulty and complexity in finding the closed-form expression of the MGF of γ_b directly using (13), PA method can be an alternative and efficient way to approximate the MGF. Its main advantage is that due to the form of the produced rational approximation, the error rates can be calculated directly using simple expressions. The PA of the MGF is a rational function of a specified order M for the denominator and L for the nominator, whose power series expansion agrees with the (M+L)-order power expansion of the MGF. Thus, the MGF of γ_b over non-i.i.d KG fading channels can be expressed as:

$$\mu_{\gamma_{b-KG}}(n) = \sum_{i=0}^{m_2-1} \frac{G_{2,2}^{2,2}[\Xi_2/\Xi_1 \mid 1-k_1-n, 1-m_1-n \mid k_2, i]}{i! \Gamma(m_1) \Gamma(k_1) \Gamma(k_2) \Xi_1^n} + \sum_{i=0}^{m_1-1} \frac{G_{2,2}^{2,2}[\Xi_1/\Xi_2 \mid 1-k_2-n, 1-m_2-n \mid k_1, i]}{i! \Gamma(m_2) \Gamma(k_2) \Gamma(k_1) \Xi_2^n}. \tag{15}$$

3.1 STATISTICS OVER MG FADING MODEL

Based on (9), the CDF of γ_b can be expressed as

$$F_{\gamma_b}(\gamma) = F_{\gamma_1}(\gamma) + F_{\gamma_2}(\gamma) - F_{\gamma_1}(\gamma) F_{\gamma_2}(\gamma), \tag{10}$$

where $F_{\gamma_i}(\gamma), i \in \{1,2\}$, is the CDF of the i^{th} -hop link. By substituting (6) into (10) and taking the derivative of (10) with respect to γ , the PDF of γ_b over non-i.i.d MG fading can be found as:

$$M_{\gamma_{b-KG}}(s) \simeq \frac{\sum_{i=0}^L c_i s^i}{1 + \sum_{i=0}^M b_i s^i} = \sum_{n=0}^{L+M} \frac{\mu_{\gamma_{b-KG}}(n)}{n!} s^n + O(s^{n+1}), \tag{14}$$

where $\mu_{\gamma_{b-KG}}(n)$ is the n th moment of γ_b , the coefficients c_i and b_i are real constants, $O(s^{n+1})$ is the remainder after truncation. In order to obtain an accurate approximation of the MGF, we assume sub-diagonal PA(M=L+1). For (13), by using [18, (03.04.26.0009.01)] and [18, (07.34.21.0011.01)], $\mu_{\gamma_{b-KG}}(n)$ can be expressed as:

4 Performance Analysis

In this section, using the previously derived closed-form expressions of the MGF over MG and KG fading channels, the average BER/SER and the diversity order of the dual-hop relaying system are obtained, respectively.

4.1 AVERAGE BER/SER

Using the MGF-based approach, we can obtain the closed-form expression of average BER/SER for the dual-hop relaying system over MG and KG fading channel. For many coherent demodulation schemes, the average SER of M-ary phase-shift keying signals (M-PSK) can be given by [17]:

$$P_{e-MPSK}(s) = \frac{1}{\pi} \int_0^{(M-1)\pi/M} M_\gamma \left(\frac{g_M}{\sin^2 \theta} \right) d\theta, \quad (16)$$

where $g_M = \sin^2(\pi/M)$. Similarly, the average SER of other modulations, such as M-ary quadrature amplitude modulation (M-QAM), can also be evaluated, which have to be neglected due to limited space in this paper.

$$|M_{\gamma_{KG}}(s)| \cong \frac{\Gamma(|m-k|)\Xi^k}{\Gamma(m)} |s|^{-k} + \frac{\Gamma(|k-m|)\Xi^m}{\Gamma(k)} |s|^{-m} + O(|s|^{-\min(k,m)-1}). \quad (18)$$

When $k = m$, (4) can be rewritten by using [18, (07.34.03. 0392.01)] as:

$$M_{\gamma_{KG}}(s) = (\Xi/s)^m U(m, 1, \Xi/s), \quad (19)$$

where $U(a, b, z)$ is the confluent hypergeometric function defined in [15]. When $s \rightarrow \infty$, by using the asymptotic series expansions of $U(a, b, z)$ in [18,

$$|M_{\gamma_{MG}}(s)| = \sum_{i=1}^N \frac{a_i C\Gamma(m)}{2\rho^m} \left(s^{-m} + \sum_{k=1}^{\infty} \binom{-m}{k} (b_i/\rho)^k s^{-(m+k)} \right) \cong \sum_{i=1}^N \frac{a_i C\Gamma(m)}{2\rho^m} |s|^{-m} + O(|s|^{-(m+1)}). \quad (21)$$

Since the values of $F_{\gamma_i}(\gamma)$ in (10) range between 0 and 1, the last two terms may be much less than their addition when $\rho \rightarrow \infty$. Hence, we can derive an

$$|M_{\gamma_{b-MG}}(s)| \cong \sum_{i=1}^N \frac{C\Gamma(m_1)a_i}{2\rho^{m_1}(s+b_i/\rho)^{m_1}} + \sum_{j=1}^N \frac{C\Gamma(m_2)a_j}{2\rho^{m_2}(s+b_j/\rho)^{m_2}}. \quad (22)$$

Similar as (21), (22) can be further approximated when $s \rightarrow \infty$, as:

$$|M_{\gamma_{b-MG}}(s)| \cong \sum_{i=1}^N \frac{C\Gamma(m_1)a_i}{2\rho^{m_1}} |s|^{-m_1} + \sum_{j=1}^N \frac{C\Gamma(m_2)a_j}{2\rho^{m_2}} |s|^{-m_2} + O(|s|^{-\min(m_1, m_2)-1}) \quad (23)$$

The above BER/SER of the dual-hop relaying system over MG and/or KG fading channels can be numerically evaluated by substituting (12) and/or (14) into (16). This can be done with some elementary numerical integration techniques.

4.2 DIVERSITY ORDER

The diversity order is the magnitude of the slope of the error probability versus SNR curve (log-log scale) in the high SNR region. The array gain measures the shift of error probability curve to the left. The diversity order and the array gain relate to the asymptotic value of the MGF near the infinity, i.e., if the MGF, $M_\gamma(s)$, can be written as

$$|M_\gamma(s)| = a |s|^{-b} + O(|s|^{-(b+1)}), \quad \text{as } s \rightarrow \infty, \quad (17)$$

where a and b are defined as the array gain and diversity order in [19], and $O(|s|^{-(b+1)})$ represents the terms of order higher than b .

When $s \rightarrow \infty$, using the asymptotic series expansions of Meijer-G function in [18, (07.34.06.0018.01)], the approximate expression of (4) when $k \neq m$ can be given as:

[18, (07.33.06.0004.01)], the approximate expression of (4) when $k = m$ can be given as

$$|M_{\gamma_{KG}}(s)| \cong \frac{2\Xi^k}{\Gamma(k)} |s|^{-k} + O(|s|^{-(k+1)}). \quad (20)$$

By using the binomial series expansion in [15, (1.110)], the approximate expression of (7) can be written as in [13]:

approximating MGF of γ_b by neglecting the product term in (10) over MG fading channel, as:

By using the same approach as (23), the approximating MGF of γ_b over KG fading channel can be expressed as (24). Based on the definition of diversity order in (17), the diversity orders of the single-hop and the dual-hop link over KG fading channel are $\min(k, m)$ and $\min(m_1, k_1, m_2, k_2)$, respectively. For MG fading

channel, they are m and $\min(m_1, m_2)$, respectively. Note that the diversity order over KG fading channel is determined by the value of m or k , not just the value of m in [12], and the diversity order over MG fading channel is determined only by the value of m as same as that over NL fading channel in [12].

$$|M_{\gamma_b-KG}(s)| \cong \begin{cases} \frac{\Gamma(|m_1-k_1|)\Xi_1^{k_1}}{\Gamma(m_1)|s|^{k_1}} + \frac{\Gamma(|k_1-m_1|)\Xi_1^{m_1}}{\Gamma(k_1)|s|^{m_1}} + \frac{\Gamma(|m_2-k_2|)\Xi_2^{k_2}}{\Gamma(m_2)|s|^{k_2}} + \frac{\Gamma(|k_2-m_2|)\Xi_2^{m_2}}{\Gamma(k_2)|s|^{m_2}} + O(|s|^{-\min(m_1, k_1, m_2, k_2)-1}) & k_1 \neq m_1 \\ & k_2 \neq m_2 \\ \frac{\Gamma(|m_1-k_1|)\Xi_1^{k_1}}{\Gamma(m_1)|s|^{k_1}} + \frac{\Gamma(|k_1-m_1|)\Xi_1^{m_1}}{\Gamma(k_1)|s|^{m_1}} + \frac{2\Xi_2^{k_2}}{\Gamma(k_2)|s|^{k_2}} + O(|s|^{-\min(m_1, k_1, k_2)-1}) & k_1 \neq m_1 \\ & k_2 = m_2 \\ \frac{\Gamma(|m_2-k_2|)\Xi_2^{k_2}}{\Gamma(m_2)|s|^{k_2}} + \frac{\Gamma(|k_2-m_2|)\Xi_2^{m_2}}{\Gamma(k_2)|s|^{m_2}} + \frac{2\Xi_1^{k_1}}{\Gamma(k_1)|s|^{k_1}} + O(|s|^{-\min(k_1, m_2, k_2)-1}) & k_1 = m_1 \\ & k_2 \neq m_2 \\ \frac{2\Xi_1^{k_1}}{\Gamma(k_1)|s|^{k_1}} + \frac{2\Xi_2^{k_2}}{\Gamma(k_2)|s|^{k_2}} + O(|s|^{-\min(k_1, k_2)-1}) & k_1 = m_1 \\ & k_2 = m_2 \end{cases} \quad (24)$$

5 Numerical and simulation results

In this section, various performance evaluation results derived by numerical and simulation techniques are presented.

Firstly, Figure 1 illustrates the BER of BPSK of single-hop link over NL, KG and MG fading, respectively, where $N=10$ for MG distribution. It can be seen from Figure1 that the performance over MG fading has almost the same as the one over NL fading. However, from Figure 1(a), it can be seen that the performance over

KG fading almost approaches to the one over NL fading as the value of k increases. From Figure 1(b), it can be seen that the former is more deviation from the latter as the value of m increases at high SNR region. Moreover, for discussing their diversity orders, we give the approximate performance over MG and KG fading channel by using (18) and (21) at high SNR, and the case of $m=1$ and $k=1$. It can be seen that the diversity order over NL and MG fading is determined by the value of m , and the diversity order over KG fading is determined by the minimum value between k and m .

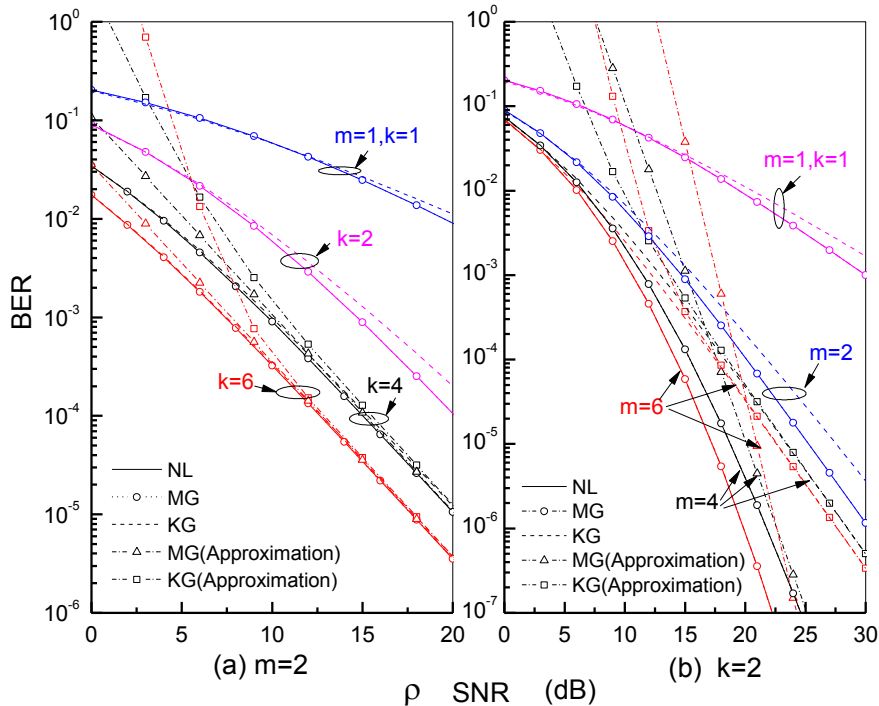


FIGURE 1 Average BER of BPSK for the single-hop link versus un-faded SNR (ρ)

In Figure 2, we give the average SER of 16PSK over i.i.d fading channels. In this case, an symmetric network geometry is assumed where R is located at the middle of a straight line between S and D, and $d_0=1$, $\varepsilon = 2$. Each

hop has the same fading parameters (k and m), $N = 10$ for MG distribution. From Figure 2, it is evident that the performance of the dual-hop link (dash line) is improved with an increase of m and/or k over MG fading.

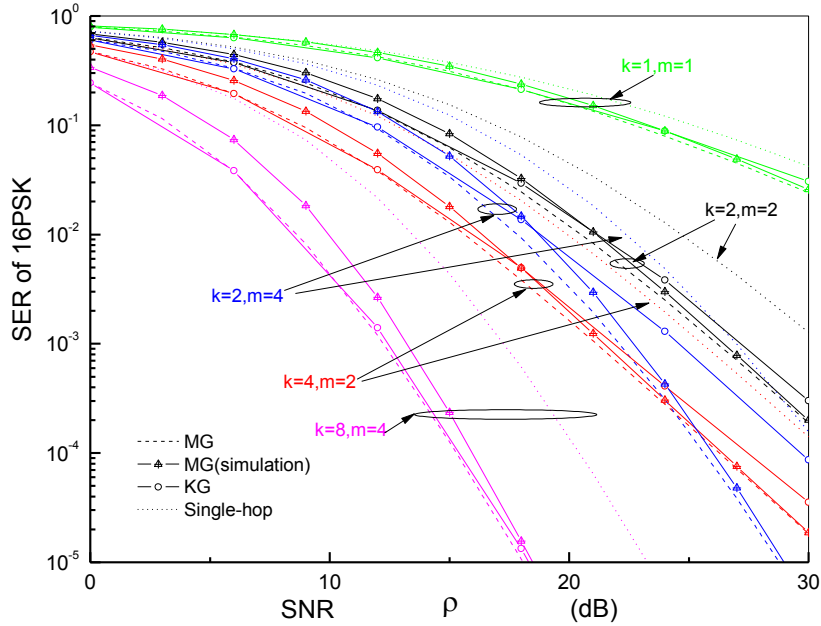


FIGURE 2 Average SER of 16PSK for the dual-hop link versus unfaded SNR (ρ) over i.i.d fading channels

They outperform the performance of the single-hop (dot line) due to the path loss reduction, and the difference become more evident as the value of m (or k) increases. For the difference between MG (dash line) and KG (line with circle mark), it can be found that they have almost similar performance when k is more than m and/or at low SNR region, but they show more deviation when k is not more than m at medium and high SNR region. Moreover, for the diversity order, it can be seen that the diversity orders over MG fading depend on

the value of m . However, the diversity orders over KG fading depend on the minimum value between k and m . At the same time, the simulation results of the dual-hop link (line with triangle mark) show agreement with the analytical results (dash line) at medium and high SNR region, and verify the mathematical accuracy.

Finally, Figure 3 shows the average SER of 16PSK of the dual-hop link over non-i.i.d MG and KG fading channels, where each hop has different channel parameters, $N=10$ for MG distribution.

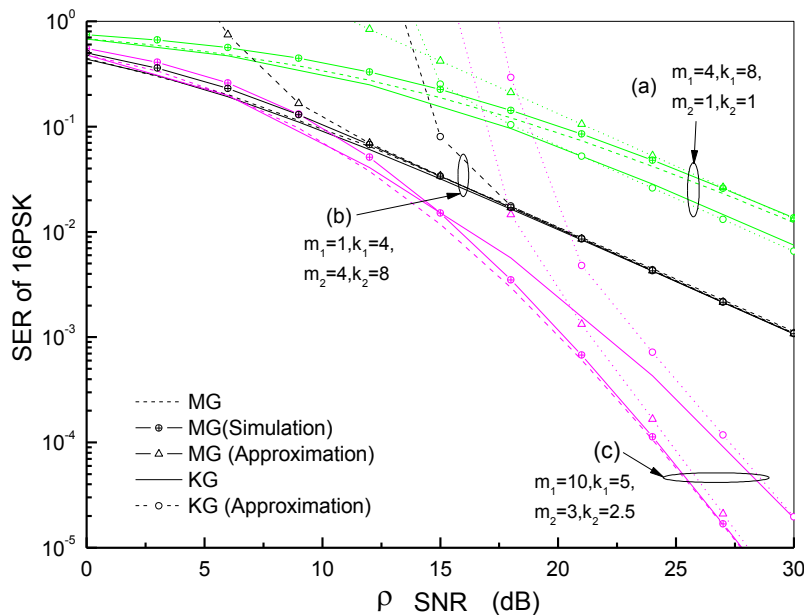


FIGURE3 Average SER of 16PSK for the dual-hop link versus unfaded SNR (ρ) over non-i.i.d fading channels

In Figure 3, their approximation results are also given by using (23) and (24). These results are similar as the ones in Figure 1 and Figure 2. From these figures, it can be seen that they show agreements between the composite NL distribution and the MG distribution if the value of N is enough large, however, only some results match well between the MG distribution and the KG distribution, for example, the case that the value of k is more than the value of m , and/or the case of low SNR region.

6 Conclusion

In this paper, the end-to-end performance of a dual-hop AF relaying system is investigated over non-i.i.d NL

fading channels by using MG distribution and KG distribution, respectively. We derived some novel closed-form expressions for the PDF and MGF of the end-to-end SNR, and evaluated some performance criterion. Numerical and simulation results are shown to verify the accuracy of the analytical results. These results show that it is more precise to approximate the NL distribution by using the MG distribution than using the KG distribution in the performance analysis of cooperative relaying systems. These works in this paper can be helpful to analyse the performance of cooperative relaying systems with co-channel interference over composite NL fading channels in the future.

References

- [1] Laneman J N, Tse D N C, Wornell G W 2004 *IEEE Transactions Information Theory* **50** 3062-80
- [2] Hasna M O, Alouini M S 2003 *IEEE Trans. Wireless Commun.* **2** 1126-31
- [3] Karagiannidis G K, Tsiftsis T A, Mallik R K 2006 *IEEE Trans. Commun.* **54** 18-22
- [4] Ikki S, Ahmed M H 2007 *IEEE Commun. Letter* **11** 334-6
- [5] Lianhai W, Jiaru L, Kai N, Zhiqiang H 2009 Performance of dual-hop transmissions with fixed gain relays over generalized-k fading channels *Proceedings of the ICC, Germany* 1-5
- [6] Efthymoglou G P, Bissias N, Aalo V A 2010 On the error rate analysis of dual-hop amplify-and-forward relaying in Generalized-K fading channels *Journal of Electrical and Computer Engineering*, 1-5 doi:10.1155/2010/584594
- [7] Bissias N, Efthymoglou G P, Aalo V A 2012 *International Journal of Electronics and Communications* **66**(1) 39-44
- [8] Peppas K P, Datsikas C K, Nistazakis H E, Tombras G S 2010 *Journal of the Franklin Institute* **347**(9) 1643-53
- [9] Datsikas C K, Peppas K P, Lazarakis F I, Tombras G S 2010 *International Journal of Electronics and Communications* **64**(11) 1094-9
- [10] Jianfei C, Lie-Liang Y, Zhangdui Z H 2010 Performance analysis of multihop wireless links over Generalized-K fading channels *IEEE Vehicular Technology Conference (Fall) Beijing* 1-5
- [11] Atapattu S, Tellambura C, Jiang H, 2010 Representation of composite fading and shadowing distributions by using mixtures of gamma distributions *Proc. IEEE Wireless Commun.* 1-5
- [12] Atapattu S, Tellambura C, Jiang H 2011 *IEEE Trans. Wireless Commun.* **10** 4193-203
- [13] Bithas P S, Sagias N C, Mathiopoulos P T, Karagiannidis G K, Rontogiannis A A 2006 *IEEE Commun. Letter* **10** 353-5
- [14] Gradshteyn I S, Ryzhik I M 2000 *Table of Integrals, Series and Products* New York: Academic
- [15] Abramowitz M, Stegun I A 1965 *Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables* Dover Publications
- [16] Abdi A, Kaveh M 1998 *Electron Letter* **34** 851-2
- [17] Simon M K, Alouini M S 2005 *Digital Communication over Fading Channels* New York: Wiley
- [18] *The Wolfram functions site* <http://functions.wolfram.com>

Author



Weijun Cheng

Current position, grades: School of Information Engineering (Minzu University of China), Associate Professor, PhD
University studies: M.Sc degree in Electronics and Control Engineering from China University of Mining and Technology, Beijing, China, in 1998, and the Ph.D. degree in Telecommunications Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2004. He was a Postdoctoral Research Fellow from 2005 to 2007 in Electronics Engineering from Peking University, Beijing, China.
Scientific interest: wireless cooperative communication theory and radio resource management.
Publications: 40 articles, 2 books

A novel task deployment approach based on graph theory for power saving

**Gaochao Xu¹, Peng Liu¹, Xiaodong Fu¹,
Yunmeng Dong¹, Jia Zhao², Yan Ding^{1*}**

¹ College of Computer Science and Technology, Jilin University, Qianjin Str. 2699, 130012 Changchun, China

² College of Computer Science and Engineering, ChangChun University of Technology, Yan'an Str. 2055, 130012 Changchun, China

Received 1 April 2014, www.tsi.lv

Abstract

With the increasing of the big datacenter, the power consumption seems to be another overhead except the equipment cost. Saving the power of big datacenter is the hotspot now. In this paper, we proposed TA-BG algorithm based on the linear weighted and graph theory to speed up the execution of tasks. Firstly, utilizing linear weighted to execute first filter to reduce the searching scope for the next research. Secondly, seeking out the hosts that can execute tasks fast based on graph theory. Finally, placing the host on the hosts selected above. The experiments indicate that TA-BG can save power of datacenter by reducing the executing time. Besides, the TA-BG even performs well on load balance.

Keywords: Cloud Data Centre, Task Allocation, Power Saving, Graph Theory

1 Introduction

Cloud computing can provide the users infrastructure, platform, and software in the form of service. Users can obtain the service as needed conveniently by the network. The cloud data centre provides on-demand compute and storage resources for the users. With the development of cloud computing, the number of big data centre is increasing. Meanwhile, the data centre needs to execute the tasks submitted by the users, and respond to the users quickly to guarantee the good user experience. Because the number of physical host is large in the cloud data centre to provide services and the amount of remaining resources are changing constantly. Therefore, how to allocate the tasks, which comes within an internal, is a research hotspot. The allocation policy will have influences on the speed of executing tasks, overhead, and load balancing etc.

Nowadays, cloud data centres often utilize the algorithm of random allocation (RA) or optimum allocation (OA) to deploy tasks to the data centres. The RA allocate the tasks to the physical hosts randomly, the advantages are getting a fast allocation and may have the lowest overhead and highest load balance in a short time. However, after a long time, the data centre can't perform well. The OA is the most used algorithm to allocate tasks in the data centre now. The algorithm selects the hosts, which have the most remaining resources to execute the tasks. The algorithm can get a fast execution, and better load balance. However, it often leads to high overheads and resource waste.

In this paper, aiming to propose an algorithm to find the allocation policy, which can make the data centre execute tasks fast, reducing the overhead of the data centre. Certainly, to make the data centre fast, low overhead in the long period, a high efficient and reasonable method to allocate the tasks to the resource pool is needed. The proposed allocation algorithm can finish the work of selecting the best physical hosts for the tasks efficiently. Especially, when used in large scale data centres, it performs well.

The rest of the paper is organized as follows. In Section 2, we present the related work about task allocation algorithms of data centres. In Section 3, present the problem the proposed algorithm solved and model of the problem. In Section 4, firstly describe the structure of the algorithm, and then state the idea of the proposed algorithm, give the detail procedure of the algorithm. In Section 5, the experimental results and analysis on CloudSim platform are given. Finally, in Section 6, we summarize the full paper and future work is put forward.

2 Related work

As far as we know, some task allocation algorithms [14, 15] are proposed and they almost stress all the goals of the task allocation. Some pay focus on the overhead of the datacenters, some stress load balance of datacenters and some aim to have a better exposed service. The algorithms that pay attention to low overhead include

*Corresponding author e-mail: dingyan11@mails.jlu.edu.cn

using the idea of constraint programming (CP) [1, 7] and the allocation based on forecast.

CP is a popular idea to solve the problem of task allocation [2, 5-7]. [2] adopts the idea of CP, thinking over a variety of constraints comes when allocate tasks. Furthermore, formulating the constraints and converting them to restricted condition of the objective function which make it convenient to get the optimal allocation policy that satisfy constraints. [3] also adopt the idea of constraint programming, it divided the task allocation into two phases, the first phase confirm the number of the host needed by the task using the idea of CP, the second phase also utilizing the CP to allocate the task to the physical host. [4] assume the traditional task allocation algorithm, which is based on forecasting, it forecast the load of the datacenter according to the history based on which they confirm the amount of the host, which should work. Then allocate the tasks to the fixed number of host which is got above. Because of the surplus host can be halt, it can reduce the overhead.

Some allocations with hope of making the load balance. Related algorithms often divided into static load balance algorithm and dynamic load balance algorithms. The static allocations [8-10] often use Round Robin, Weighed Round Robin, Weighted Least-Connection Scheduling and Least-Connection Scheduling. [8] Wei Qun adopt the Weighted Least-Connection Scheduling, that is, show the host's performance with different weight and allocate the tasks to the hosts which has the lowest ratio of amount and weight. These static algorithm only utilize some static information, they cannot adapt to the dynamic load variation of the datacenter efficiently. Dynamic load balance algorithm [11, 12] is a classic combined optimization problem and is proved to be NP hard. In addition, it often incurs extra communication overhead during the procedure of balancing the load dynamically. Nowadays, the best algorithm to solve dynamic load balance seems to be greedy algorithm. In [13], Lau combined heavy-load preferred and light-load preferred and proposed an adaptive load-dispatching algorithm, reducing the communication overhead during balancing the load.

3 Proposed problem and its formulation

In the data centre, a group of tasks need to allocate to the host in the data centre to deal with. However, the number of host in the data centre is large, and this will certainly lead to some different kinds of allocation policy. Furthermore, what the policy is adopted will bring the data centre diverse cost. Therefore, to find a high-efficient, lower-cost task allocation policy is necessary.

We now formulate the problem of allocating n tasks onto m physical hosts. Its solution can be represented by an n dimension of solution vector, each element of which denotes the target host of the tasks. We defined as follows: H is a set of m available physical hosts denoted by $H(h, t) = \{h_1, h_2, h_3, \dots, h_m\}$, available at time t , each

host have the cost denoted by $c(s)$ ($s \in \{1, 2, 3, \dots, m\}$). R represents a set of resource, including CPU, memory, disk resources. $u_{rj}(r \in R, j \in H)$ represents the amount of resource r in the host j . Besides, A is a set of the tasks which come to the data centre within Δt , denoted by $A = \{a_1, a_2, a_3, \dots, a_n\}$. $t_{ri}(r \in R, i \in A)$ represents how many r resources the task i need. Define the 0-1 variable x_{ij} , when task i is allocated to the host j , $x_{ij} = 1$, otherwise $x_{ij} = 0$. Similarly, define the 0-1 variable y_j , when the host j have tasks $y_j = 1$, otherwise $y_j = 0$. Above all, the target function can be present as follows:

$$\min \sum_{j \in H} y_j c_j, \tag{1}$$

$$\text{s.t. } \sum_{i \in A} \sum_{j \in H} x_{ij} = n$$

$$\text{s.t. } t_{ri} \leq u_{ri}, (r \in R, x_{ij} = 1)$$

$$c(s) = \alpha \cdot \text{CPU}(s) + \beta \cdot \text{Mem}(s) + d(s), \tag{2}$$

where the $\text{CPU}(s)$ denotes the power consumption of host s , the unit is kw/h , it can be got from the time t 's function.

The $\text{mem}(s)$ denotes the power consumption of the memory of host s , the unit is kw/h , it is the function of one variable time t .

The $d(s)$ denotes other cost, including depreciation cost, using cost and so on. It is determined by the data centre manager, each host has different value d . α, β denote the unit price of the power, the unit is *yuan* per kw/h .

The first constraint makes sure that all the tasks are allocated to the hosts and the second constraint confirms that host i must have enough resources when the task j is allocated to the host i . Finally, we will get the solution vector denoted by $TH, TH = \{s_1, s_2, s_3, \dots, s_n\} (s_k \in H)$.

4 Description of the proposed allocation algorithm TA-BG

4.1 THE PROPOSED SYSTEM ARCHITECTURE

Figure 1 illustrates the system architecture of the data centre, in which it utilizes the TA-BG algorithm to allocate tasks. Task allocation controller consists of four parts: the part of task collection, task allocation, host information base and the core of the architecture, which is used for computing the allocation policy using TA-BG algorithm. The part of task collection is responsible for collecting the tasks from the internet, and cache them. t seconds later, it sends the set of tasks to the TA-BG controller. The TA-BG controller will compute the allocation policy combined the tasks and the information of hosts. At last, the optimal policy, which computed by TA-BG will send to the part of task allocation, the task is allocated to the target physical hosts.

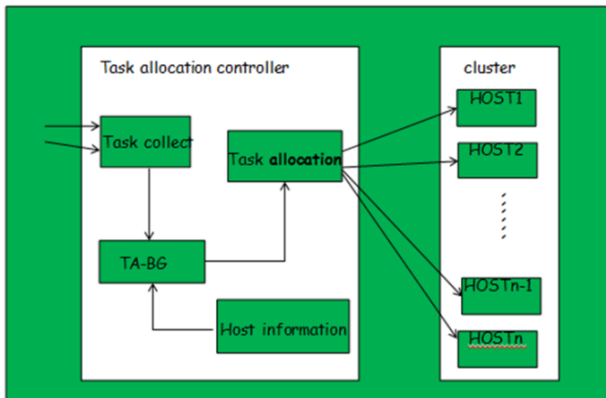


FIGURE 1 System Architecture of TA-BG

4.2 THE SOLUTION REPRESENTATION

By using the above system architecture, we can get the optimal allocation policy. In detail, since there are n tasks during one time of allocation, so the solution should be denoted by n dimensional vector, that is $TH = \{s_1, s_2, s_3, \dots, s_n\}$ ($s_k \in H$). Every component of solution vector should be computed by the TA-BG algorithm, denote that which host the task is allocated to. For example, $s_1 = 2$ represent that task 1 is allocated to the host 2.

4.3 PROPOSED TA-BG ALGORITHM

The proposed TA-BG algorithm mainly utilizes the graph theory to solve the problem of task allocation of data centre. It can obtain an allocation policy with high efficient and low overhead. TA-BG is comprised of three steps. Firstly, filter hosts preliminary, filtering the hosts which even cannot execute the task with minimum resource in the purpose of reducing the number of next selection. Secondly, filter in the set that is got above. Here we use the concept of degree in the graph theory, regarding each host as a node and two nodes have edges only when they communicate each other. By doing so, the physical cluster will abstract to an undirected graph. Finally sort the hosts from large to small based on degree, then allocate the tasks from the beginning of the ordinal sequence.

4.3.1 Filter the hosts with few resources

The step is to filter the host that cannot execute any task roughly and quickly, to reduce the searching scope for the next step. So we take linear weighted to denote the resource of host j , that is the sum of CPU, memory etc. denoted by variable R_j^h .

$$R_j^h = \alpha \cdot CPU_j + \beta \cdot mem_j + \gamma \cdot disk_j, \tag{3}$$

s.t. $\alpha + \beta + \gamma = 1$.

where CPU_j denotes the surplus resources of host j , mem_j denotes the surplus memory of host j . α, β, γ is the weight

factor, representing the importance of every variable and their sum must be 1.

To compare, we need to abstract the task demand to the similar variable. Denoted by R_i^t :

$$R_i^t = \alpha \cdot CPU_i + \beta \cdot mem_i + \gamma \cdot disk_i. \tag{4}$$

After the formulation, it's convenient to compare, we need not to compare the every component of the vector respectively. Besides, we can utilize heap sort to find the task with minimum resources request, and then compare with each host's surplus resources, weeding out the hosts that do not satisfy the condition $R_i^t \leq R_j^h$. From this step, we can get the set $H1$ preliminary.

4.3.2 Filter the hosts with better connectivity

Here we need to abstract the data centre to an undirected graph so that we can find the better connectivity host conveniently, so regarding each host in the cluster to a node and two hosts have an edge only when they have communicated each other. After the formulation, the connectivity of the host in the data centre can be represented by the degree of the nodes in the undirected graph. According to the experience and analysis, the host with better connectivity may have better performance. What's more, the better connectivity hosts must be in the centre of the cluster that is near from controller of data centre. Moreover, because the host must communicate with controller when execute tasks. So using the better connectivity host to execute the task will certainly reduce the communication overhead. Now CPUs seems to be fast enough, but communication delay still have not enough progress because restrict of physical links. By doing so, we can reduce the communication overhead skilfully.

Sort the host in the $H1$ from large to small based on degree, and dispatch the tasks to the host with larger degree. We hold the principle that the task should be allocated to the host with lager degree as possible, the surplus hosts can be halt to reduce overhead.

5 Evaluation

In this section, we have experimentally verified the performance of the proposed TA-BG approach. The verification includes three aspects: the speed of executing tasks, the load balance and power consumption. Simultaneously, we have taken other two-allocation algorithm to conduct experiment, which also includes the same above three aspects to show the performance of TA-BG obviously. The first comparison experiment adopts RA to allocate the tasks, and the second comparison experiment takes OA to allocate the tasks. In experiment, we allocated 100 tasks to the data centre using the three-allocation algorithm separately and observe their performance.

To simulate a dynamic cloud data centre, we adopted the cloud simulation tools Cloudsim to conduct experiments. During the simulation period, Cloudsim can calculate power consumption by the getPower() method and view the condition of load. Cloudsim can create all kinds of entities and delete or add entities on the running machine dynamically, which is the reason why we choose the Cloudsim to conduct experiments.

5.1 EXPERIMENTAL SCENARIOS

In the Cloudsim platform, establishing a resource pool with 100 physical hosts and each host has the different number of resources. Simulating 100 task requests, all of them have disparate resource request like CPU, memory etc. By using the Cloudsim, we can get the information of the cloud data centre to further analysis.

5.2 COMPARISON OF RA, OA AND TA-BG IN EXECUTING SPEED

The experiment is designed for verifying the executing speed of the TA-BG algorithm. We firstly prepare 500 tasks with the same size. Allocating the 500 tasks to the cloud datacenters in succession, and recording the time every 100 tasks are finished. We conclude the followed bar graph according to the experiment:

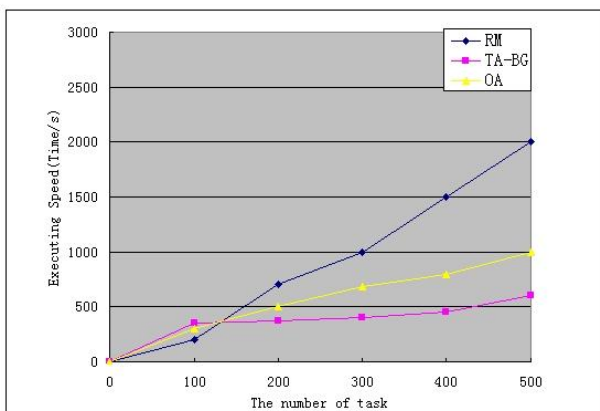


FIGURE 2 Comparison of Executing Speed

From the Figure 2 we can conclude that: the data centre executes the tasks fastest and be the most stable when taking TA-BG approach to allocate the tasks.

5.3 COMPARISON OF RA, OA AND TA-BG IN LOAD BALANCE

In the experiment scenario, we evaluate the TA-BG algorithm in load balance. We allocated 100 tasks to the host of datacenters that adopt three different allocation algorithms respectively. We recorded the condition of load on each host every ten minutes, and then calculated the variance. Here we only utilize the using rate of CPUs to represent the host's load. u_i denotes the current using rate of host i , m denotes the number of physical host. The

average utilization of all hosts' CPUs is denoted by the formula (5) the degree of load balance denoted by the formula (6):

$$\bar{u} = \frac{\sum_{i=1}^m u_i}{m}, \tag{5}$$

$$B = \sqrt{\frac{1}{m} * \sum_{i=1}^m (u_i - \bar{u})^2}. \tag{6}$$

We get the curve chart according to the recode:

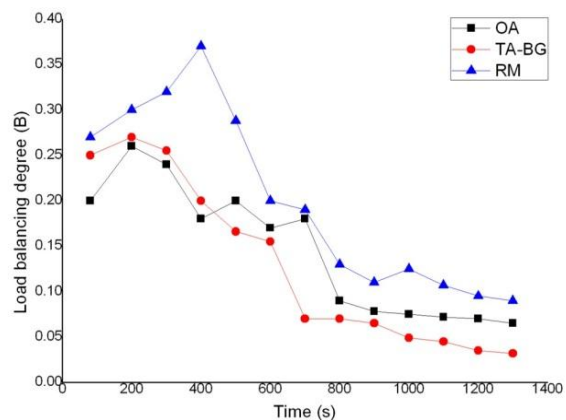


FIGURE 3 Comparison of load balancing degree

Small variance means the using rate of all host's CPUs are closed. So small variance means the load is more balanced. Namely, the value B is the smaller the better.

From the Figure 3, when using the traditional RA algorithm the load balance is worse than other two algorithm from begging to end. At the beginning, the datacenter, which adopt OA algorithm is more balanced than the one adopt TA-BG algorithm. However, when $t > 420s$ the datacenter with TA-BG algorithm is better than them with OA algorithm on load balance. Therefore, we can conclude that the proposed TA-BG have good performance on load balance especially on the long-running datacenter.

5.4 COMPARISON OF RA, OA AND TA-BG IN POWER CONSUMPTION

In this experimental scenario, verifying the efficiency and availability of TA-BG in power saving. We also need three experiments, allocating 100 tasks to the different datacenters, the first one adopt RA to allocate tasks, the second adopt OA and the third use proposed TA-BG. We tested the power consumption every ten minutes, and draw the Figure 4.

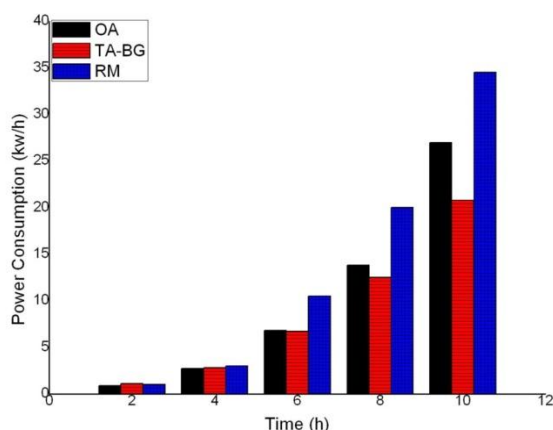


FIGURE 4 Comparison of energy consumption

As illustrated in Figure 4. When $t > 6h$ the power consumption is still lower than other two datacenters. And with the increase of the load in the cloud data center, the incremental power consumption of TA-BG is less than that of OA and RA obviously. And when $t > 6h$ the power consumption is still lower than other two datacenters.

6 Conclusion and future work

In this paper, an allocation algorithm of datacenter is proposed, and we give its idea, implementation and

References

- [1] Müller T 1994 *Interactivity in Constraint Programming* DFKI of documentation series, German research center for artificial intelligence (dfki), stuh.
- [2] Dhyani K, Gualandi S, Cremonesi P 2010 A Constraint Programming Approach for the Service Consolidation Problem *CPAIOR 2010 LNCS 6140* 97–101
- [3] Hermenier F, Lorca X, Menaud J-M 2009 Entropy: a Consolidation Manager for Clusters *VEE'09, March Washington, DC, USA* 11–3
- [4] Rolia J, Andrzejak A, Arlitt M F 2003 Automating enterprise application placement in resource utilities *DSOM 2003 LNCS* Eds Brunner M, Keller **2867** 118–29
- [5] Hermenier F, Demassey S, Xavier L 2011 *CP 2011 LNCS 6876* 27–41
- [6] Sirdey R, Carlier J, Kerivin H, Dritan N 2007 *European Journal of Operational Research* **183** 546–63
- [7] Fukunaga A S 2009 Search Spaces for Min-Perturbation Repair. In proceeding of: *Principles and Practice of Constraint Programming - CP 2009 15th International Conference, CP 2009 Lisbon, Portugal, September 20-24*
- [8] Wei Qun, Xu Guangli, Li Yuling 2011 Research on duster and load balance based on LINUX virtual server *Information Computing and Applications. Berlin: Springer Heidelberg* 169-76
- [9] Song S, Lv T, Chen X 2014 A Static Load Balancing algorithm for Future Internet *TELKOMNIKA Indonesian Journal of Electrical Engineering* **12**(6)
- [10] Chen Wei, Zhang Yufang, Xiong Zhongyang 2010 Research and realization of the load balancing algorithm for heterogeneous cluster with dynamic feedback *Journal of Chongqing University* **33**(2) 2-14
- [11] Willebeek-LeMair M H, Reeves A P 1993 Strategies for dynamic load balancing on highly parallel computers *IEEE Transactions on Parallel and Distributed Systems* **4**(9) 979-93
- [12] You T, Li W, Fang Z 2014 Performance Evaluation of Dynamic Load Balancing Algorithms *TELKOMNIKA Indonesian Journal of Electrical Engineering* **12**(4)
- [13] Lau S M, Lu Q, Leung K S 2006 Adaptive load distribution algorithms for heterogeneous distributed systems with multiple task classes *Journal of Parallel and Distributed Computing* **66**(2) 163-80
- [14] Shen Kaiji, Zheng Xiaoying 2012 Fair multi-node multi-resource allocation and task scheduling in datacenter, *Cloud Computing Congress (APCloudCC), 2012 IEEE Asia Pacific* 59 – 63
- [15] Pedram Massoud 2012 Energy-Efficient Datacenters *IEEE Trans. on Computer Aided Design* **31**(10) 1465 – 84

evaluation. It is based on linear weighted and graph theory. The first step, utilize linear weighted to filter the hosts with small resources. The second step, use graph theory to find the hosts that can finish the tasks as soon as possible. The third step, allocate the tasks to the hosts which are chosen from above steps. Finally, we conduct three teams of experiments, selecting OA and RA as the comparison experiment, to verify the performance of the TA-BG algorithm. They separately verify the advantages of TA-BG on executing speed, load balance and power saving. From above experiment, we can conclude that the TA-BG algorithm has better performance on several aspects than other allocation algorithm, especially fitting for the long-running datacenters. It is an available and efficient allocation algorithm for the real physical datacenter.

Aiming to further improve the performance of TA-BG, we hope to apply some other algorithm to allocate the task to the hosts selected by the first two steps instead of allocating them randomly. By doing this, it seems to be more predominant on all kinds of performance mentioned above like speed, load balance and power consumption.

Authors	
	<p>Gaochao Xu, born in 1966, Wuhan</p> <p>Current position, grades: Changchun, Professor</p> <p>University studies: BS, MS and PhD on College of Computer science and Technology of Jilin University in 1988, 1991 and 1995.</p> <p>Scientific interest: Cloud Computing, Mobile Cloud Computing, distributed system, grid computing, cloud computing, Internet of things, information security, software testing and software reliability assessment</p> <p>Publications: SCI 10</p> <p>Experience: he is the professor and PhD supervisor of College of Computer Science and Technology, Jilin University, China. As a person in charge or a principal participant, Dr Gaochao Xu has finished more than 10 national, provincial and ministerial level research projects of China.</p>
	<p>Peng Liu, born in 1990, Jixi of Heilongjiang province of China</p> <p>Current position, grades: a postgraduate student of the college of computer science and technology of Jilin University</p> <p>University studies: the study of computer science at Daqing Normal University in 2009 and got his bachelor degree in 2013.</p> <p>Scientific interest: Virtualization, Cloud Computing, Mobile Cloud Computing, SDN</p>
	<p>Xiaodong Fu, Changchun</p> <p>Current position, grades: Senior engineer in the College of Computer Science and Technology, Jilin University of China</p> <p>University studies: B.Sc. degree from Jilin University</p> <p>Scientific interest: Distributed System, Grid Computing, Cloud Computing, Internet Things, etc</p> <p>Publications: 14 research articles</p>
	<p>Yunmeng Dong, born in 1989, Yushu of Jilin province of China in, ChangChun</p> <p>Current position, grades: Changchun, Master</p> <p>University studies: studies of computer science at Changchun University of Technology in 2008, bachelor degree in 2012. Now he is a postgraduate candidate of the college of computer science and technology of Jilin University Virtualization</p> <p>Scientific interest: Cloud Computing, Mobile Cloud Computing Virtualization, Cloud Computing, Mobile Cloud Computing</p> <p>Publications: EI 1</p> <p>Experience: Yunmeng Dong. He began the. His main research interests include distributed system, cloud computing and virtualization technology</p>
	<p>Jia Zhao, born in 1982, Changchun of Jilin province of China, Changchun</p> <p>Current position, grades: PhD candidate of the college of computer science and technology of Jilin University</p> <p>University studies: master degree in 2008</p> <p>Scientific interest: Virtualization, Cloud Computing, Mobile Cloud Computing</p> <p>Publications: SCI 4</p> <p>Experience: distributed system, cloud computing, network technology. He has participated in several projects.</p>
	<p>Yan Ding, born in 1988, Yichun of Heilongjiang province of China, Tieli</p> <p>Current position, grades: Changchun, Master</p> <p>University studies: bachelor degree in 2011. Now he is a postgraduate candidate of the college of computer science and technology of Jilin University.</p> <p>Scientific interest: Virtualization, Cloud Computing, Mobile Cloud Computing</p> <p>Publications: SCI 1</p> <p>Experience: He began the study of computer science at Jilin University in 2007 and got his degree</p>

Image enlargement based on the hyperbolic Coons interpolation

Juncheng Li*

Department of Mathematics and Econometrics, Hunan University of Humanities, Science and Technology, Dixing Road, Loudi, 417000, China

Received 1 March 2014, www.tsi.lv

Abstract

A method for image enlargement, making use of the hyperbolic Coons interpolation surface with shape parameters, is investigated in this paper. As a non-polynomial model, the hyperbolic Coons interpolation surface can represent the image better than the general interpolation methods. By altering the values of the shape parameters, the effects of image enlargement can be adjusted until achieving the satisfactory results. Experimental results show that the effects of image enlargement making use of the hyperbolic Coons interpolation surface are better than the general interpolation methods.

Keywords: image enlargement, hyperbolic Coons interpolation surface, shape parameters

1 Introduction

As an important part of image processing, image enlargement has been widely applied in different areas. Because a two-dimensional static grey image can be represented by a continuous function of two variables, surface interpolation methods are generally used to deal with image enlargement problems. The general interpolation methods for image enlargement include nearest interpolation, bilinear interpolation, bicubic interpolation, cubic spline interpolation, and so on [1, 2], which used polynomial models to construct image interpolation surfaces. However, the polynomials can only reflect the gradual change of data but not the mutability of data, the quality of the target images require to be further improved. Therefore, the uses of non-polynomial models would be better choices for constructing image interpolation surfaces.

Generally, curves and surfaces are established based on polynomial functions in Computer Aided Geometric Design (CAGD), in particular curve and surface design. Non-polynomial models, such as trigonometric or hyperbolic curves and surfaces, have gained very much interest in recent years within CAGD. Zhang constructed the C-Bézier curve and surface in the space $\{1, t, \sin t, \cos t\}$ [3]. Mainar and Chen defined the C-Bézier curves of higher order in the space $\{1, t, \dots, t^{k-3}, \cos t, \sin t\}$ [4, 5]. Wang constructed the non-uniform algebraic trigonometric B-splines in the space $\{1, t, \dots, t^{k-3}, \cos t, \sin t\}$ [6]. Han presented a cubic trigonometric Bézier curve with two shape parameters in the space $\{1, \sin t, \cos t, \sin^2 t\}$ [7]. Li constructed a class of quasi-cubic trigonometric curves in the space $\{1, \sin t, \cos t, \sin^2 t\}$ [8]. Liu presented another kind of trigonometric Bézier curve with two shape parameters in the space $\{1, \sin t, \cos t,$

$\sin^2 t\}$ [9]. Yan discussed a class of algebraic-trigonometric blended splines in the space $\{1, t, \sin t, \cos t, \sin^2 t, \sin^3 t, \cos^3 t\}$ [10]. Lü presented the hyperbolic polynomial B-splines in the space $\{\sinh t, \cosh t, t^{k-3}, t^{k-4}, \dots, t, 1\}$ [11], and Li extended these hyperbolic splines to the case of non-uniform knot vector [12]. Liu studied a kind of hyperbolic polynomial uniform B-spline surface with shape parameter in depth [13]. Li presented the hyperbolic polynomial Ferguson curve and Coons surface with shape parameters, which analogous to the corresponding cubic polynomial Ferguson curve and bicubic Coons surface [14].

Although the trigonometric and hyperbolic curves and surfaces have been widely studied, but up to now, the uses of the trigonometric or hyperbolic models for constructing interpolation surfaces in image processing have rarely been studied. The main purpose of this work is to investigate the use of the hyperbolic Coons interpolation surface with two shape parameters [14] to deal with image enlargement problems.

The present work is organized as follows. In Section 2, the basic principle of image zooming based on surface interpolation is given. In Section 3, the hyperbolic Coons interpolation surface is reviewed, and then image zooming making use of the hyperbolic Coons interpolation surface is investigated. In Section 4, the experimental results and discussion are provided. A short conclusion is given in Section 5.

2 The basic principle of image enlargement based on surface interpolation

The basic principle of image enlargement based on surface interpolation has two steps. The first step is to construct a continuous surface interpolating all the pixel

*Corresponding author e-mail: lijuncheng82@126.com

points of the original image, and the second step is to obtain the target image by resampling the interpolation surface according to the enlargement ratio.

Given a grey image $I(x, y)$ with the size of $M \times N$, and $g_{i,j}$ ($i=0,1,\dots,M-1$; $j=0,1,\dots,N-1$) is the grey value with row i and column j according to the point (i, j) in the pixel plane. For enlarging the original image $I(x, y)$ to the target image $I'(x, y)$ with the size of $M_1 \times N_1$, firstly the continuous interpolation surface $F(x, y)$ ($0 \leq x \leq M-1$, $0 \leq y \leq N-1$) of the original image is constructed by some interpolation methods, then the target image can be obtained by uniform selecting $M_1 \times N_1$ pixel points in the interpolation surface $F(x, y)$. Therefore, the grey value g'_{i_1, j_1} with row i_1 and column j_1 of the target image $I'(x, y)$ satisfies

$$g'_{i_1, j_1} = F\left(\frac{M}{M_1}i_1, \frac{N}{N_1}j_1\right), \tag{1}$$

$$\mathbf{p}(u, v) = \begin{bmatrix} F_0(u; \alpha) & F_1(u; \alpha) & G_0(u; \alpha) & G_1(u; \alpha) \end{bmatrix} \begin{bmatrix} \mathbf{p}_{00} & \mathbf{p}_{01} & \mathbf{p}_{v00} & \mathbf{p}_{v01} \\ \mathbf{p}_{10} & \mathbf{p}_{11} & \mathbf{p}_{v10} & \mathbf{p}_{v11} \\ \mathbf{p}_{u00} & \mathbf{p}_{u01} & \mathbf{p}_{uv00} & \mathbf{p}_{uv01} \\ \mathbf{p}_{u10} & \mathbf{p}_{u11} & \mathbf{p}_{uv10} & \mathbf{p}_{uv11} \end{bmatrix} \begin{bmatrix} F_0(v; \beta) \\ F_1(v; \beta) \\ G_0(v; \beta) \\ G_1(v; \beta) \end{bmatrix}, \tag{2}$$

where \mathbf{p}_{00} , \mathbf{p}_{01} , \mathbf{p}_{10} and \mathbf{p}_{11} are point vectors in the four corner points; \mathbf{p}_{u00} , \mathbf{p}_{u01} , \mathbf{p}_{u10} and \mathbf{p}_{u11} are directional derivatives of direction u ; \mathbf{p}_{v00} , \mathbf{p}_{v01} , \mathbf{p}_{v10} and \mathbf{p}_{v11} are directional derivatives of direction v ; \mathbf{p}_{uv00} , \mathbf{p}_{uv01} , \mathbf{p}_{uv10}

where $i_1 = 0, 1, \dots, M_1 - 1$; $j_1 = 0, 1, \dots, N_1 - 1$.

According to the process of image enlargement, the key step is to construct the continuous image interpolation surface. In the following discussion, the hyperbolic polynomial Coons interpolation surface with two shape parameters [14] will be adopted to deal with image enlargement problems.

3 Image enlargement making use of the hyperbolic coons interpolation surface

3.1 THE HYPERBOLIC COONS INTERPOLATION SURFACE

For two arbitrary real number α and β , $\alpha, \beta > 0$, $0 \leq u \leq \alpha$, $0 \leq v \leq \beta$, the hyperbolic Coons interpolation surface [14] can be expressed as:

and \mathbf{p}_{uv11} are mixed directional derivatives; $F_i(t; \delta)$, $G_i(t; \delta)$ ($i=0,1$; $t=u, v$; $\delta=\alpha, \beta$) are the hyperbolic Hermite basis functions defined as:

$$\begin{cases} F_0(t; \delta) = \frac{1}{2-2C+\delta S} \left((1-C+\delta S) - St + S \sinh t + (1-C) \cosh t \right) \\ F_1(t; \delta) = \frac{1}{2-2C+\delta S} \left((1-C) + St - S \sinh t - (1-C) \cosh t \right) \\ G_0(t; \delta) = \frac{1}{2-2C+\delta S} \left(-(S-\delta C) + (1-C)t + (1-C+\delta S) \sinh t + (S-\delta C) \cosh t \right) \\ G_1(t; \delta) = \frac{1}{2-2C+\delta S} \left(-(\delta-S) + (1-C)t - (1-C) \sinh t + (\delta-S) \cosh t \right) \end{cases}, \tag{3}$$

where $S := \sinh \delta$, $C := \cosh \delta$.

In order to use the hyperbolic Coons interpolation surface defined as Equation (2) to deal with image enlargement problems, we redefine the surface on the

fixed interval $[0,1] \times [0,1]$. Let $s = \frac{u}{\alpha}$ and $t = \frac{v}{\beta}$, the hyperbolic Hermite basis functions can be redefined as follows:

$$\begin{cases} F_0(s) = F_0(u; \alpha), & F_1(s) = F_1(u; \alpha), & G_0(s) = \frac{1}{\alpha} G_0(u; \alpha), & G_1(s) = \frac{1}{\alpha} G_1(u; \alpha) \\ F_0(t) = F_0(v; \beta), & F_1(t) = F_1(v; \beta), & G_0(t) = \frac{1}{\beta} G_0(v; \beta), & G_1(t) = \frac{1}{\beta} G_1(v; \beta) \end{cases}, \tag{4}$$

According to Equation (2) and Equation (4), the function of the hyperbolic Coons interpolation surface can be expressed as follows:

$$f(s,t) = \begin{bmatrix} F_0(s) & F_1(s) & G_0(s) & G_1(s) \end{bmatrix} \begin{bmatrix} f_{00} & f_{01} & f_{t00} & f_{t01} \\ f_{10} & f_{11} & f_{t10} & f_{t11} \\ f_{s00} & f_{s01} & f_{sr00} & f_{sr01} \\ f_{s10} & f_{s11} & f_{sr10} & f_{sr11} \end{bmatrix} \begin{bmatrix} F_0(t) \\ F_1(t) \\ G_0(t) \\ G_1(t) \end{bmatrix}, \tag{5}$$

where $0 \leq s \leq 1$ and $0 \leq t \leq 1$.

interpolation surface defined as Equation (5) has the following properties at the endpoints:

By simple deduction, the hyperbolic Coons

$$\begin{cases} f(0,0)=f_{00}, & f(0,1)=f_{01}, & f(1,0)=f_{10}, & f(1,1)=f_{11} \\ \frac{\partial f(0,0)}{\partial s}=f_{s00}, & \frac{\partial f(0,1)}{\partial s}=f_{s01}, & \frac{\partial f(1,0)}{\partial s}=f_{s10}, & \frac{\partial f(1,1)}{\partial s}=f_{s11} \\ \frac{\partial f(0,0)}{\partial t}=f_{t00}, & \frac{\partial f(0,1)}{\partial t}=f_{t01}, & \frac{\partial f(1,0)}{\partial t}=f_{t10}, & \frac{\partial f(1,1)}{\partial t}=f_{t11} \\ \frac{\partial^2 f(0,0)}{\partial s \partial t}=f_{sr00}, & \frac{\partial^2 f(0,1)}{\partial s \partial t}=f_{sr01}, & \frac{\partial^2 f(1,0)}{\partial s \partial t}=f_{sr10}, & \frac{\partial^2 f(1,1)}{\partial s \partial t}=f_{sr11} \end{cases} \tag{6}$$

Remark 1: Equation (6) shows that the hyperbolic Coons interpolation surface defined as Equation (5) has the same interpolation properties to the general bicubic polynomial Coons surface. However, the shape of the general bicubic polynomial Coons surface can not be adjusted when the boundary conditions are kept unchanged, while the shape of the hyperbolic Coons interpolation surface defined as Equation (5) can be easily adjusted by altering the value of the two shape parameters α and β . Given the boundary conditions

The effects on shape of the hyperbolic Coons interpolation surface by altering the value of α and β are shown in Figure 1, where (a) is generated by setting $\alpha=1$ and $\beta=5$, (b) is generated by setting $\alpha=2$ and $\beta=10$, (c) is generated by setting $\alpha=5$ and $\beta=1$, (d) is generated by setting $\alpha=10$ and $\beta=2$.

are:

$$\begin{bmatrix} f_{00} & f_{01} & f_{t00} & f_{t01} \\ f_{10} & f_{11} & f_{t10} & f_{t11} \\ f_{s00} & f_{s01} & f_{sr00} & f_{sr01} \\ f_{s10} & f_{s11} & f_{sr10} & f_{sr11} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

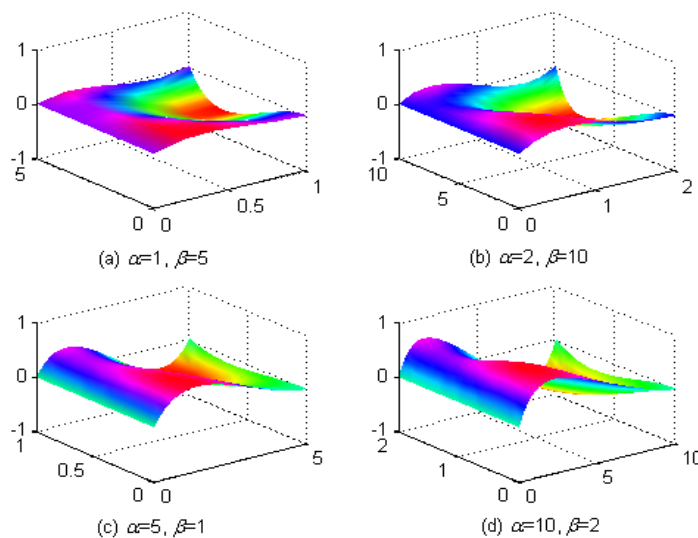


FIGURE 1 The hyperbolic Coons interpolation surfaces with different shape parameters

3.2 IMAGE ENLARGEMENT METHOD

Considering the region: $\Delta_{ij} = \{(x, y) | i \leq x \leq i+1, j \leq y \leq j+1\}$, $(i=0,1,\dots,M-2; j=0,1,\dots,N-2)$ in the pixel plane of

the original image $I(x, y)$.

Let $s = x - i$ and $t = y - j$, according to Equation (5) and Equation (6), the image interpolation surface in Δ_{ij} can be expressed as follows:

$$f_{ij}(s, t) = [F_0(s) \quad F_1(s) \quad G_0(s) \quad G_1(s)] \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} F_0(t) \\ F_1(t) \\ G_0(t) \\ G_1(t) \end{bmatrix}, \tag{7}$$

where a_{mn} ($m, n = 1, 2, 3, 4$) satisfy:

$$\begin{cases} a_{11} = g_{i,j}, & a_{12} = g_{i,j+1}, & a_{13} = g_{i,j+1} - g_{i,j}, & a_{14} = g_{i,j+2} - g_{i,j+1}, \\ a_{21} = g_{i+1,j}, & a_{22} = g_{i+1,j+1}, & a_{23} = g_{i+1,j+1} - g_{i+1,j}, & a_{24} = g_{i+1,j+2} - g_{i+1,j+1}, \\ a_{31} = g_{i+1,j} - g_{i,j}, & a_{32} = g_{i+1,j+1} - g_{i,j+1}, & a_{33} = a_{34} = a_{43} = a_{44} = 0, \\ a_{41} = g_{i+2,j} - g_{i+1,j}, & a_{42} = g_{i+2,j+1} - g_{i+1,j+1}, & i = 0, 1, \dots, M-2; j = 0, 1, \dots, N-2 \end{cases}$$

and set $g_{i,N} = 2g_{i,N-1} - g_{i,N-2}$, $g_{M,j} = 2g_{M-1,j} - g_{M-2,j}$.

Remark 2: In order to get Equation (7), we use forward difference quotient to instead derivative and take the torsional vector of each corner point as zero. From Equation (7), the entire interpolation surface $F(x, y)$ of the original image $I(x, y)$ can be expressed as:

$$F(x, y) = f_{i,j}(s, t), \tag{8}$$

where $s = x - i$, $t = y - j$, $i \leq x \leq i+1$, $j \leq y \leq j+1$.

According to Equation (8), the entire image interpolation surface $F(x, y)$ ($0 \leq x \leq M-1, 0 \leq y \leq N-1$) is connected by $(M-1) \times (N-1)$ pieces of hyperbolic Coons interpolation surfaces and satisfies C^1 continuous.

For enlarging $I(x, y)$ to the target image $I'(x, y)$

with the size of $M_1 \times N_1$, let $x = \frac{M}{M_1} i_1$ and $y = \frac{N}{N_1} j_1$,

according to Equation (8), then $i = \lfloor x \rfloor$, $j = \lfloor y \rfloor$, where $\lfloor \cdot \rfloor$ is the round down functions.

Taking $s = x - i$ and $t = y - j$ into Equation (7), the grey values of the target image $I'(x, y)$ can be obtained by $g'_{i_1, j_1} = f_{i,j}(s, t)$, where $i_1 = 0, 1, \dots, M_1 - 1$; $j_1 = 0, 1, \dots, N_1 - 1$.

Remark 3: In contrast with the general interpolation methods, the use of hyperbolic Coons interpolation surface for constructing the image interpolation surface has the following advantages:

- 1) As a non-polynomial model, the hyperbolic polynomial not only can reflect the gradual change of data, but also can reflect the mutability of data;
- 2) The hyperbolic Coons interpolation surfaces directly interpolate the pixel points of the image, which avoiding solving a matrix system;
- 3) The hyperbolic Coons interpolation surface has two shape parameters α and β , the local characteristics of the target image can be proper adjusted by altering the two shape parameters.

4 Experimental results and discussions

The six images shown in Figure 2 are used as examples. In Figure 2, the size of (a) to (c) is 300×400 (8 bit) and the size of (d) to (f) is 256×256 (8 bit).

The effect of image enlargement is quantitatively compared by the PSNR (dB), which is defined as follows,

$$\text{PSNR} = 10 \times \log \frac{255^2}{\text{MSE}},$$

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - I'(i, j)]^2, \tag{9}$$

where $I(i, j)$ and $I'(i, j)$ is the original image and the target image respectively, $M \times N$ represents the size of the original image.

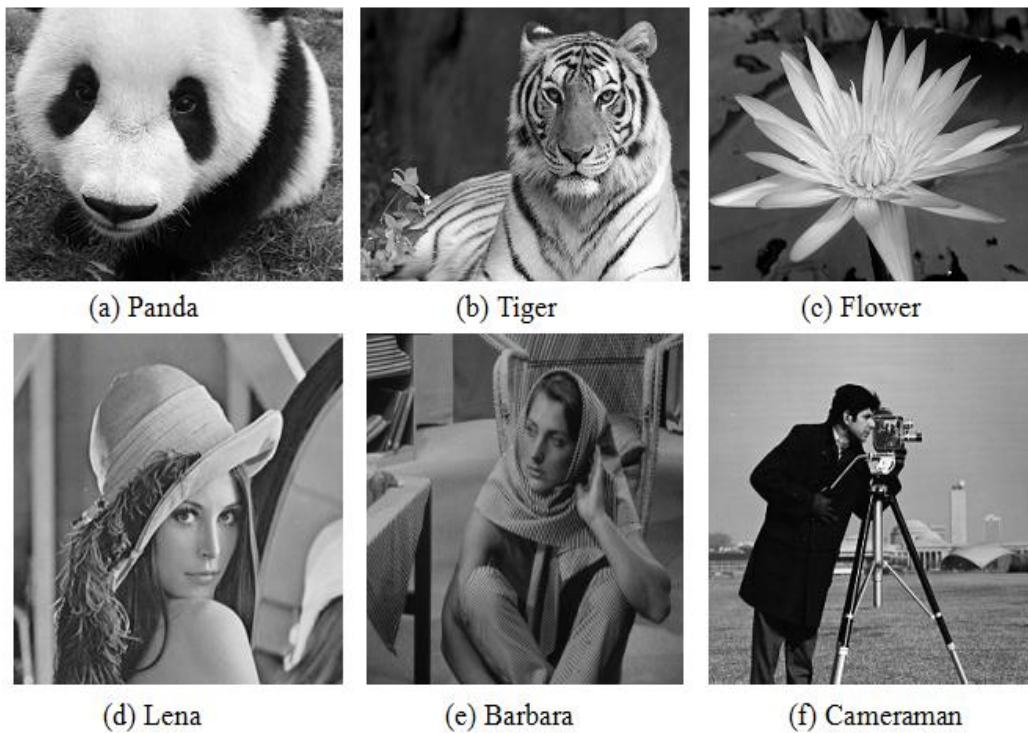


FIGURE 2 Original images for experiment

4.1 INFLUENCE OF THE SHAPE PARAMETERS ON IMAGE ENLARGEMENT

Firstly, in order to illustrate the effectiveness of the proposed method, we enlarge part of the Panda image (right part with the size of 100×60) shown in Figure 2

with 2 times and 3 times respectively. The enlarged images are shown in Figure 3, where the shape parameters of the image interpolation surface are set for $\alpha = \beta = 300$.

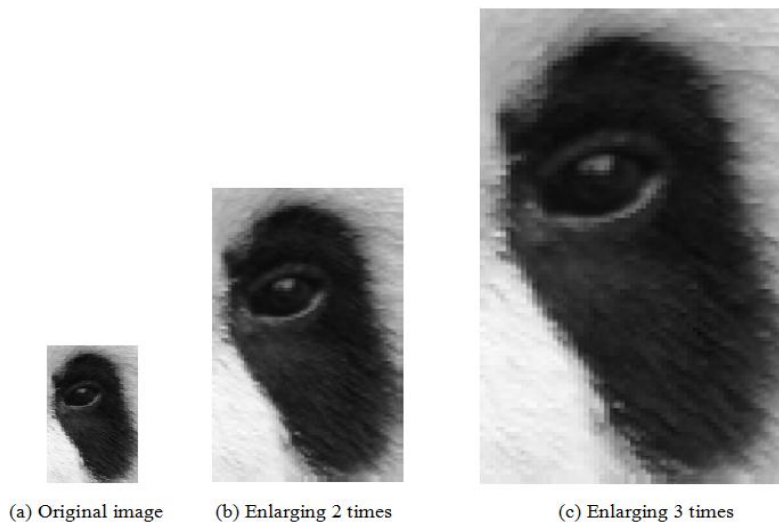


FIGURE 3 Experimental results of part of the Panda image

Figure 3 shows that the enlarged image does not appear obviously mosaic phenomenon and block phenomenon, and keep the edge information well. Even when the magnification is higher, the enlarged images still have good visual effects.

Secondly, we enlarge part of the Tiger image (face

part with the size of 100×100) shown in Figure 2 with 3 times. The enlarged images, when the two shape parameters are of different values, are shown in Figure 4.

The edge detection of the images in Figure 4 using the Prewitt operator (the threshold is set for 0.1) are shown in Figure 5.

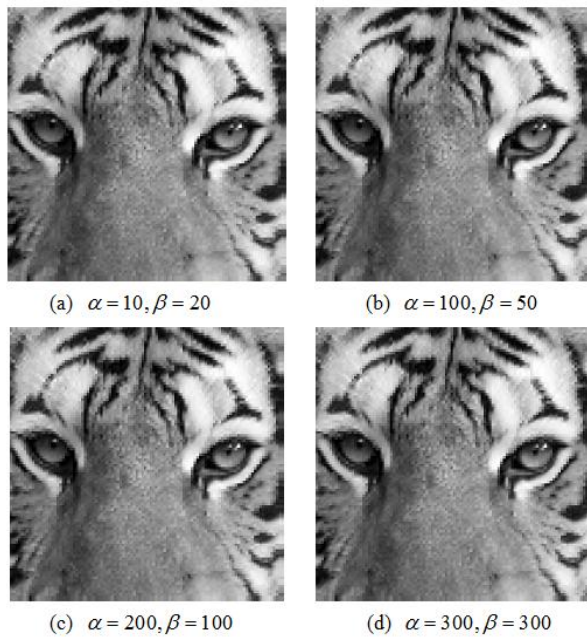


FIGURE 4 Experimental results of part of the Tiger image

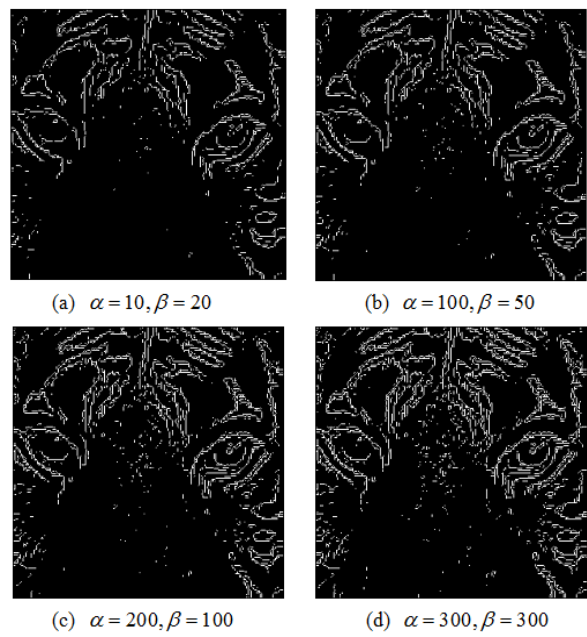


FIGURE 5 Edge detection of the images in Figure 4

Figure 4 and Figure 5 show that, when increasing the values of shape parameters α and β , the enlarged images would have gradually clear edge profile and boundary.

Remark 4: Through a large number of experiments on different images, we find that the shape parameter α and β have great influence on image enlargement. In general, the values of the shape parameters are greater, the better the effects of image enlargement. However, when the values of the shape parameters are too large, the enlarged images will produce certain distortion. So, in practical applications, the values of the shape parameters can be selected appropriately at first, if the target images are not satisfactory, we can modify the values of the shape parameters until obtaining satisfactory target

images.

4.2 COMPARISON OF THE PROPOSED METHOD WITH THE GENERAL METHODS

In order to compare the effects of image enlargement between the proposed method and the general interpolation methods, we firstly reduce the Flower image shown in Figure 2 under four minification using the bilinear interpolation, then enlarge the contractible images under four times again using the proposed methods (the shape parameters are set $\alpha = \beta = 300$) and the general interpolation methods respectively. Experimental results of the Flower image using different methods are shown in Figure 6.

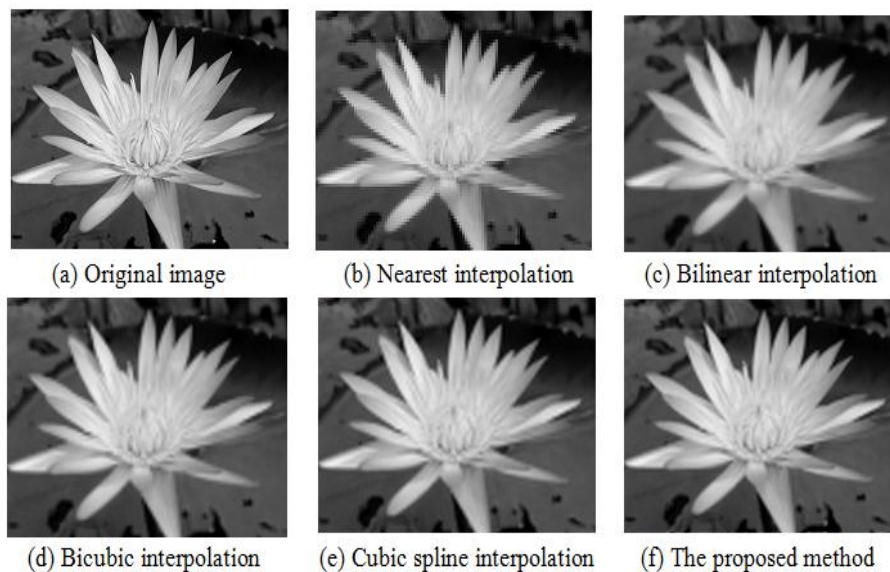


FIGURE 6 Experimental results of Flower image

Figure 6 shows that the proposed method can get better visual effect than the four general interpolation methods.

For quantitatively comparing the effects of image enlargement between the proposed method and the general interpolation methods, we firstly reduce the three original images shown in Figure 2(d)~(f) under four minification using bilinear interpolation, then enlarge the contractible images under four times again using the proposed methods (the shape parameters are set $\alpha = \beta = 300$) and the general interpolation methods respectively. The PSNR is computed to evaluate every method, which is listed in Table 1.

TABLE 1 The PSNR of different interpolation methods

Method	PSNR (dB)		
	Lena	Barbara	Cameraman
Nearest interpolation	32.4601	31.0211	33.0331
Bilinear interpolation	32.7356	31.2772	32.7071
Bicubic interpolation	33.2165	31.6113	32.9770
Cubic spline interpolation	33.3041	31.6939	32.9835
The proposed method	33.8866	32.0870	33.7275

References

- [1] Castleman K-R 1995 *Digital Image Processing*. New Jersey: Prentice Hall
- [2] Gonzalez R-C, Woods R-E 2005 *Digital Image Processing Using Matlab* New Jersey: Prentice Hall
- [3] Zhang J W 1999 *Graphical Models and Image Processing* 61(1) 2-15
- [4] Mainar E, Pena J M 2002 *Computer Aided Geometric Design* 19(4) 291-5
- [5] Chen Q Y, Wang G Z A 2003 *Computer Aided Geometric Design* 20(1) 29-39
- [6] Wang G Z, Chen Q Y, Zhou M H 2004 *Computer Aided Geometric Design* 21(2) 193-205
- [7] Han X-A 2009 *Applied Mathematics Letters* 22(2) 226-31
- [8] Li J C, Zhao D B, Li B J, Chen G H A 2010 *Journal of Information and Computational Science* 7(13) 2847-54
- [9] Liu H Y, Li L, Zhang D M 2011 *Journal of Information and Computational Science* 8(7) 1217-23
- [10] Yan L L, Liang J F 2011 *Journal of Computational and Applied Mathematics* 235(6) 1713-29
- [11] Lü Y G, Wang G Z, Yang X N 2002 *Computer Aided Geometric Design* 19(6) 379-93
- [12] Li Y J, Wang G Z 2005 Two Kinds of B-Basis of the Algebraic Hyperbolic Space *Journal of Zhejiang University SCIENCE A* 6(7) 750-9 (in Chinese)
- [13] Liu X M, Xu, W X, Guan Y, Shang Y Y 2010 *Graphical Models* 72(1) 1-6
- [14] Li J C, Yang L, Zhong Y-E, Xie C 2012 *Communications in Computer and Information Science* 289 359-67

Author



Juncheng Li

Current position, grades: Hunan University of Humanities, Science and Technology (China); Lecturer

Scientific interest: Computer aided geometric design and image processing; Information and Computational Science

A fast top-down visual attention method to accelerate template matching

Yiping Shen, Shuxiao Li, Chengfei Zhu*, Hongxing Chang

Institute of Automation Chinese Academy of Sciences, Beijing, China, 95 Zhongguancun East Road, Beijing, China

Received 1 March 2014, www.tsi.lv

Abstract

This paper presents a fast top-down visual attention method to downsize the search space of template matching. Such a method first generates patterns representing the local structures, and then calculates the pattern distributions representing the template and its surroundings. From here two separate operations are performed: the "pattern weight" is first introduced, which describes how well a certain pattern is correlated to the template, and then weights of all patterns are calculated for later reference. This is the "off-line" operation, and in comparison the "on-line" operation only calculates the pattern of each pixel, whose weights can be indexed conveniently from the off-line results. With all pixels' pattern weights calculated, the weight image is ready, from which we can extract the region of interest for subsequent matching. Experiments showed that our method obtained at least 6.21 times speed-ups over the state-of-the-art methods with little or no loss in performance.

Keywords: template matching, visual attention, top-down attention, saliency, region of interest

1 Introduction

Template matching (TM) is defined as searching for a sub-window (referred as candidate in the rest of this paper) that is most similar to a given template in a larger reference image. The similarity is usually measured as a cost function, e.g., product Cross Correlation (CC) [1], Normalized Cross Correlation (NCC) [1, 2], Zero-mean Normalized Cross Correlation (ZNCC) [1, 3], Sum of Absolute Differences (SAD) [4], Sum of Squared Difference (SSD) [5,6], Hamming Distance [7]. NCC and ZNCC are widely used due to their robustness to linear brightness variations [3].

The original algorithm of TM needs to search the entire space to get the best match, which is a time consuming procedure and thus limits its application in real-time environments not to mention on devices with limited computational resources. Much work has been explored to accelerate the computation of TM, which can be categorized into two aspects [2]. One aspect is to find an efficient representation of the template which enables fast computation of the cost function, e.g., Fast Fourier Transform (FFT) [1], Walsh-Hadamard transform [5], and Haar-like binary features [2]. Other techniques aim to reduce the search space or early prune the computation where the best match unlikely locates. For example, lower bound-based methods [5, 8] were used to accelerate the computation of SSD, while upper bound-based methods [3, 9-11] were used for NCC. Bound-based methods were more efficient, yet the order in which the candidates were examined affected the run time of the algorithms. A dual-bound method proposed in [6] obtained the best possible runtime by using a

priority queue to determine an optimal ordering for examining the candidates. These efforts obtained high computation reduction. However, the run time of bound-based methods is data-dependent and may have no advantages over full space searching methods in the worst case.

Visual attention helps humans to fast focus on the information of interest when dealing with a huge mass of information [12, 13]. This property encourages researchers to bring it to machine vision systems. A considerable amount of research in cognitive science and computer vision has been conducted to understand and model visual attention mechanisms. Most of the research are concentrated on bottom-up attention (also called stimuli-driven) and relative models are built to analyse, which parts of the image attract human's attention in free viewing (for reviews please refer to [12, 14, 15]). However, the importance of the top-down (or task-driven) modulation have been emphasized in recent years [12, 16-18], and the integration of bottom-up saliency and top-down modulation models have been widely explored in [12, 16-19]. These models first computed the saliency maps based on the colour, intensity, and orientation features. Then top-down modulation was realized either by increasing the saliency on the expected location or increasing the weights of some specific features. Advances in visual attention are beneficial for solving some challenging problems in computer vision, e.g. object detection [16, 18, 19], tracking [20]. Nevertheless, some of these models involve time-consuming procedures, and current top-down models are mostly based on the bottom-up stimuli, which cannot deal with the situations in which the object

*Corresponding author e-mail: chengfei.zhu@ia.ac.cn

does not generate strong enough stimuli.

In this paper, a fast top-down visual attention method is proposed to reduce the search space of TM. The method consists of two parts: (1) a ROI is extracted based on the proposed top-down visual attention model; (2) ZNCC-based template matching is performed at each candidate of the ROI to get the final match. In the first part, the proposed method represents the local structure by patterns and builds pattern distributions for the template and the background, respectively. Note that, we can use a representative image containing the template (see Figure 1) or a set of images (see in Section 4.3) to empirically evaluate the pattern distribution for the background. If a representative image is used, we artificially warp the image and build pattern distributions with these warped images to obtain small scale invariance and in-plane rotation invariance. Then, pattern weights are calculated off-line by enhancing the template (referred as the specific object) patterns while suppressing the distracting background patterns simultaneously. These weights indicate how well the patterns are correlated to the specific object. For the on-line process, we only need to calculate the pattern for each pixel in the reference image and get the corresponding pattern weight by indexing in the learned model. This is the generation of the weight image. Then, the average weight of each candidate is computed by the integral image [21] and the one with highest average weight is the centre of the region of interest (ROI) which is extracted for subsequent template matching. The term "top-down" is referred because the generation of the

weight image is controlled by the top-down knowledge, i.e. the appearance of the specific object showed in the template. Experiments show that our method obtains 30.90, 6.21, 24.08, 2.97 times speed-ups over a sequential implementation of FFTs [22], a state-of-the-art ZNCC-based method named Two-stage extended-mode Partial Computation Elimination (TPCE) [11], a state-of-the-art SAD-based method named Partial Distortion Elimination (PDE) [4] and the highly optimized implementation based on FFT in openCV (called HFFT for short, see in <http://opencv.org>), respectively with little or no lose in performance. The advantage and novelty of our method mainly include:

- Comparing to bound-based methods, the run time of the proposed method is data independent.
- We propose a top-down visual attention model to downsize the search space. In this model patterns represent local structures and the pattern weight describes how well a certain pattern is correlated to the template. Using patterns as the stimuli and pattern weights as the strength of stimuli, the top-down control is realized by setting pattern weights learned off-line.

This paper is organized as follows. Section 2 is an introduction of ZNCC-based TM. Section 3 describes the details of the proposed method. In section 4, we verify the method with experiments and compare it with four fast TM algorithms: FFTs [22], TPCE [11], PDE [4], and HFFT. Section 5 gives the conclusion of this paper.

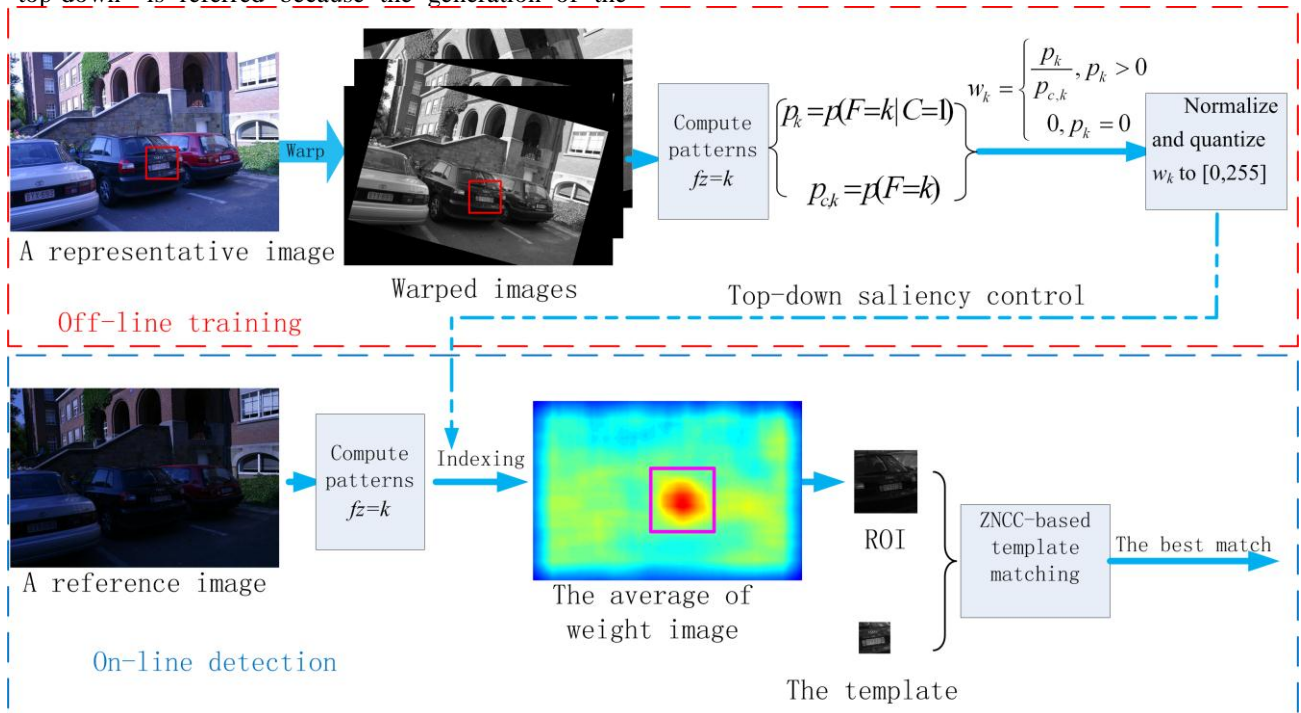


FIGURE 1 The flowchart of the proposed method. The red rectangle denotes the specify object to be detected (i.e. the template)

2 Template Matching Using ZNCC

Let I and T denote the reference image and the template,

respectively. The size of I is $M \times N$ pixels, while the size of T is $m \times n$ pixels, where $m \leq M$ and $n \leq N$. The similarity between T and I at location (x, y) can be given by:

$$ZNCC(x, y) = \frac{\sum_{j=1}^m \sum_{i=1}^n [I(x+i, y+j) - \mu(x, y)] \cdot [T(i, j) - \mu(T)]}{\sqrt{\sum_{j=1}^m \sum_{i=1}^n [I(x+i, y+j) - \mu(x, y)]^2} \cdot \sqrt{\sum_{j=1}^m \sum_{i=1}^n [T(i, j) - \mu(T)]^2}}, \quad (1)$$

where

$$\mu(x, y) = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n I(x+i, y+j), \mu(T) = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n T(i, j). \quad (2)$$

The original full searching algorithm needs to scan the whole reference image and computes the ZNCC values for all candidates. Therefore, the total computation includes mnJ additions and mnJ multiplications, where $J=(M-m+1) \times (N-n+1)$ is the number of candidates. It can be reduced to $6MN \log_2(MN)$ additions and $6MN \log_2(MN)$ multiplications by using FFT [1].

3 The Proposed Top-Down Visual Attention Method

In this section, we first introduce the calculation of local structural patterns. Then, the proposed visual attention model is established by estimating and analysing pattern distributions for the template and the background using a representative image. At last, we describe the detection procedure from reference images based on the acquired visual attention model.

3.1 LOCAL STRUCTURAL PATTERN REPRESENTATION

Intensity, colour and orientation have been commonly used in visual attention computational models [12]. In this study, we use binary strings as an efficient representation of patterns and use patterns in our attention model. This idea comes from the Local Binary Patterns (LBs) [23], which own two advantages. First and most importantly, the feature space of LBs is a finite set, which enables us to establish a table to save the properties of patterns. Thus, once the properties (i.e. pattern weights) have been learned from the representative image off-line, we can get the weight of a certain pattern in the reference image by indexing. Secondly, LBs are more robust to illumination changes [23] than intensity and colour features, and more efficient than orientation features which often involve convolutions with Gabor filters.

LBs, first introduced by Ojala et al. [23], encode the pixel-wise information in an image, and have been widely used in texture classification [24] and face recognition because of its simplicity, efficiency, grayscale invariance and satisfactory discrimination [25]. LBs describe the relationship between the centre z_c and its P neighbours z_0, z_1, \dots, z_{P-1} (see Figure 2(a)). Formally,

$$LBP_{P,r} = \sum_{i=0}^{P-1} s(I(z_i) - I(z_c)) 2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (3)$$

where $I(z_i)$ denotes the grayscale value at pixel z_i . A threshold t is used as follows [25] ($t=3$) to increase the robustness in flat areas:

$$LBP_{P,r} = \sum_{i=0}^{P-1} s(I(z_i) - I(z_c) + t) 2^i. \quad (4)$$

In this study, we make some changes to the sampling points as follows: 16 points are sampled around the centre similar to the DAISY configuration [26] as shown in Figure 2(b). Two rings are used to make the local structural pattern more distinctive. Six points are sampled equally on the inner ring while ten points are sampled equally on the outer ring. The radius of the inner ring is r , while that of outer ring is $2r$. Experiments showed that setting $r=5$ can obtain the best performance. We use a Gaussian weighted sum of the grayscale value in the neighborhood instead of the grayscale value at pixel z_c and $z_i, i = 0, 1, \dots, 15$ to deal with local distortions. The weighted sum is realized by a convolution with a 3×3 Gaussian kernel $[1 \ 2 \ 1]^T [1 \ 2 \ 1]$, where $[\]^T$ is matrix transposition.

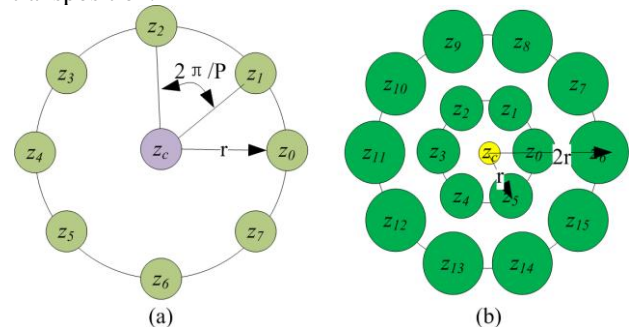


FIGURE 2 (a) The centre pixel z_c and its eight neighbours on the ring of r for $P=8$; (b) The configuration of DAISY_LBP. The small yellow circle denotes the centre; the large green circles denote 16 sampling points and the size of these circles corresponds to the smoothing range to deal with local distortions

The convolution only needs four additions and two multiplications for each pixel. Three convolutions with this kernel are used for sampling points on the outer ring, while two and one for points on the inner ring and the center point, respectively. Using (4), we get a 16-bit binary pattern $f (f \in [0,65535])$ termed as DAISY_LBP.

3.2 THE PROPOSED TOP-DOWN VISUAL ATTENTION MODEL

The basic insight of our model is that a pattern f gets

more saliency thus rewarding a higher weight if it takes place more frequently in the target than in background. Inspired by this insight and the saliency using natural image statistics model (SUN) [27] which performed well in predicting people's fixations in free viewing, we set up our top-down visual attention model. Let $C=1$ denote a

point belonging to the target, $C=0$ denote that of the background, L denote the location of a pixel, and F denote the pattern of a pixel. Assuming that patterns and locations are independent, and conditional independent for given $C=1$, the saliency s_z can be defined as:

$$s_z = p(C=1 | F=f_z, L=l_z) = \frac{p(F=f_z | C=1)}{p(F=f_z)} p(C=1 | L=l_z), \quad (5)$$

where f_z is the pattern at pixel z . Since we have no priors about the location of the target, $p(C=1 | L=z)$ can be ignored in (5). So (5) can be rewritten as:

$$s_z = p(F=f_z | C=1) / p(F=f_z), \quad (6)$$

Using (6), we need to evaluate the pattern distributions for both the target and the background. The resulting saliency thus enhances the patterns of the target while it suppresses the patterns of distracting background. Note that useless target patterns are also suppressed if the background activates the same patterns more frequently.

In our work, the pattern distribution for the target is evaluated using the template, while the distribution for the background is estimated using a representative image or a set of images. Let $p_k, k=0,1,\dots,65535$ be the probability of $f_z=k$ in the target (the numerator in (6)), $p_{c,k}$ be the probability in the background (the denominator in (6)), then the weight w_k of the pattern $k=f_z$ can be

calculated as:

$$w_k = \begin{cases} p_k / p_{c,k}, & p_k > 0 \\ 0, & p_k = 0 \end{cases}. \quad (7)$$

In this way, w_k implies the top-down control to the generation of the weight image.

We compute the whole set of $w_k, k=0,1,\dots,65535$ for a given template, and save them in a table during the off-line training phase. We also artificially warp the representative image to obtain small scale (0.85,1.15) and in-plane rotation(-15°,15°) robustness. 7 scale bins and 7 in-plane rotation bins are used in steps of 0.05 and 5°, respectively, yielding 49 warped images. At last, $w_k, k=0,1,\dots,65535$ is normalized and quantized to [0,255], which can be saved in a byte. The training phase is summarized in Table 1.

TABLE 1 Algorithms for off-line training and on-line detection

The off-line training algorithm	
Input: a representative image I_{rep} and a template.	Output: the top-down visual attention model $W=\{w_k, k=0,1,\dots,65535\}$.
(1) Warp I_{rep} using scale and in-plane rotation transform to get 49 images: $I_{rep,1}, I_{rep,2}, \dots, I_{rep,49}$. (2) For $i=1,2,\dots,49$, compute DAISY_LBP at every pixel in $I_{rep,i}$. Compute the histograms of patterns in the template $hist_t$ and the representative image $hist_r$. (3) For $k=0,1,\dots,65535$, compute p_k and $p_{c,k}$ from $hist_t$ and $hist_r$, and calculate w_k according to Equ. (7). (4) Normalize and quantize w_k to [0,255].	
The on-line detection algorithm	
Input: the reference image I_{ref}, and model $W=\{w_k, k=0,1,\dots,65535\}$	Output: best matching position and score.
(1) Compute DAISY_LBP at every pixel in image I_{ref} . (2) Assign w_{f_z} to pixel z to generate the weight image. (3) A sliding window is run across the weight image to get the location (x_{opt}, y_{opt}) with maximum average weight by the integral image. (4) Extract the ROI. (5) HFFT is performed within the ROI to yield the final match.	

3.3 FAST DETECTION FROM REFERENCE IMAGES BASED ON THE SALIENCY MODEL

For the on-line detection phase, we first get the DAISY_LBP f_z for each pixel z in the reference image. Then, a weight image is generated by assigning w_{f_z} to pixel z . Let the size of template be $m \times n$. A sliding window of $m \times n$ is used to get the average weight of each candidate, which can be accelerated by the integral

image. The candidate with the maximum average weight is considered as the centre of ROI, denoted by (x_{opt}, y_{opt}) . The size of ROI is decided according to the experimental results of the Euclidean distance between the ground truth and (x_{opt}, y_{opt}) , which will be discussed in section 4.1. Finally, HFFT is performed within the ROI to yield the final match. The on-line detection phase is summarized in Table 1.

Therefore, the computation of the on-line phase includes three convolutions with a 3×3 kernel ($12MN$ additions and $6MN$ multiplications), the computation of patterns ($16MN$ comparisons), the integral image ($4MN$ additions), the average weight ($4MN$ additions and MN comparisons) and TM within ROI ($6WH \log_2(WH)$ additions and $6WH \log_2(WH)$ multiplications with the size of ROI $W \times H$ pixels). The proposed method eliminates $J - J_{ROI}$ candidates with an overhead of $20MN$ additions, $17MN$ comparisons and $6MN$ multiplications. Here J and J_{ROI} denote the numbers of candidates in the reference image and ROI, respectively. In comparison, FFT needs $6MN \log_2(MN)$ additions and $6MN \log_2(MN)$ multiplications.

4 Experimental results

4.1 EXPERIMENTS ON IMAGES WITH GAUSSIAN NOISE

Dataset. Forty images with size 640×480 are randomly chosen from MIT database (<http://people.csail.mit.edu/torralba/images/>), which is mainly concerned with indoor and urban scenes (see Figure 4). Five different levels of Gaussian noise with peak signal-to-noise ratio (PSNR) values of 27, 24, 21, 18, and 15 are added to each image of the dataset, respectively. Two template sizes 50×50 and 100×100 are used, and for each size 10 not-too-smooth templates are randomly chosen from each image. Therefore, there are 4000 matches in total (a match is defined as the

most similar candidate found in a reference image).

Results. We evaluate the performance of the proposed method with different Gaussian noises as well as different configuration of P sample points described in Section 3.1. For $P = 8$, the original LBs are employed; for $P = 12$, the configuration of two rings with six points equally sampled on each ring is performed; and for $P = 16$, the configuration is depicted in Figure 2(b). We compute the Euclidean distance d between the ground truth and the centre of the ROI (x_{opt}, y_{opt}) , and draw the curve of $\#(d < x) / Total$ to x for each noise level as depicted in Figure 3. Here, $\#(d < x)$ denotes the number of matches with $d < x$ and $Total$ ($Total = 40 \times 10$) denotes total matches in the experiment with the same noise level and P . We can see that (x_{opt}, y_{opt}) is closer to the ground truth with larger P . We do not investigate larger P than 16 (e.g. 32) because it needs too much memory for the weight tables. Thus, the suggested value of P is 16 for our method. When $P = 16$, more than 75% and 89.5% of the total matches can be found whose distance is less than 10 pixels on sizes 100×100 and 50×50 , respectively, and more than 98.25% and 99% of the total matches whose distance is lesser than 50 pixels on sizes 100×100 and 50×50 , respectively. The size of ROI is set to be $(m+99) \times (n+99)$ for the following experiments according to this experiment, where $m \times n$ is the size of the template. Therefore, the number of candidates in ROI is $100 \times 100 = 10000$. Note that a smaller size of ROI contains less candidates in ROI thus leads to less computation. However, it may miss the most promising candidate.

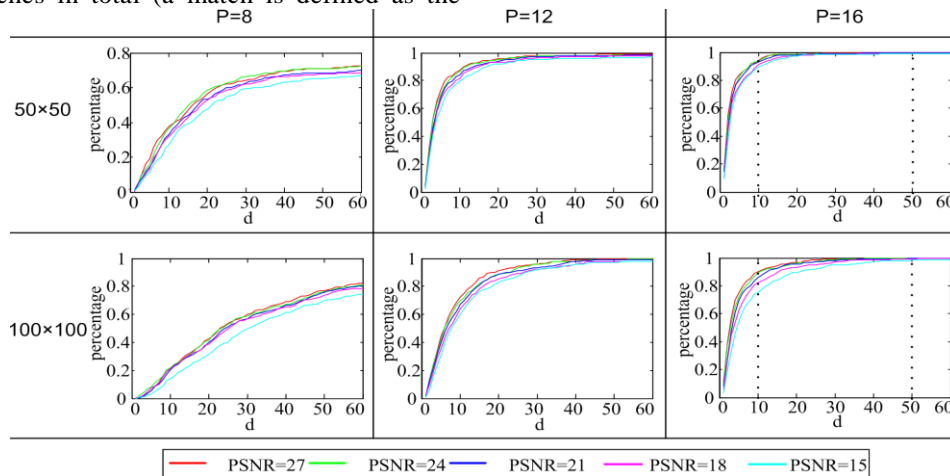


FIGURE 3 The results with Gaussian noise. The X-axis corresponds to the Euclidean distance d between the groundtruth and the center of extracted ROI (x_{opt}, y_{opt}) . The Y-axis equals to $\#(d < x) / Total$, where $\#(d < x)$ denotes the number of matches with $d < x$ and $Total$ ($Total = 40 \times 10$) denotes total matches in the experiment with the same noise level and P

4.2 EXPERIMENTS ON IMAGES WITH TRANSFORMATIONS

In this section, we compare our method with FFTs [22], TPCE [11], PDE [4] and HFFT. Demos for FFTs, TPCE and PDE are available at <http://cvlab.lums.edu.pk/pce>. The parameters of TPCE are set according to [11]. All algorithms are based on ZNCC except for PDE, which is a full search

equivalent SAD-based algorithm. Five transformations were evaluated similar to [7]: small in-plane rotation, small scale changes, illumination changes, blur, and JPEG compression. All algorithms are implemented in C++ and run on an Intel Core2 Duo CPU E4400 2.00 GHz/2G RAM computer.

Dataset. The dataset is from OX database (<http://www.robots.ox.ac.uk/~vgg/research/affine/>). We use three groups of images (see Figure 4), which are

designed to test the robustness to illumination (Leuven), blur (Bikes), and Jpeg compression (Ubc). Six images in each group, we choose the first as the representative image and run the algorithms in the other five images. To evaluate the robustness to small geometrical changes, we create two data sets for scale and rotation changes. For small scale changes, Graffiti is warped to generate 10 images with scale randomly chosen in $[-0.85, 1.15]$, and for in-plane rotation changes, Boat is rotated to yield 10 images with rotation angles in $[-15^\circ, 15^\circ]$. Therefore, we have five groups to evaluate these algorithms under the five transformations. For each group, five template sizes (32×32 , 50×50 , 64×64 , 100×100 , 128×128) are used and 40 templates are randomly chosen for each template size, yielding 7000 ($5 \times (40 \times 5 \times 3 + 40 \times 10 \times 2)$) matches. Templates with a standard deviation smaller than 60 are skipped to avoid the flat regions such as the blue sky in Ubc. Note that the templates are extracted from the representative image.

Results. Let " ROI_R ", " $FINAL_R$ " denote the results of ROI extraction and the final results of proposed method, respectively. We can easily obtain the location of the best match according to the true homography between the representative image and the reference image. The Euclidean distances d between the ground truth and the results by these algorithms are computed. We regard a match as a correct match if d is smaller than five pixels (meaning that the intersection of the detection and the ground truth exceeds 84% of the ground truth). The detection rate is defined as the number of correct matches with respect to the number of total matches. The speed-ups over FFTs in run time is defined similar to [11]. Results are showed in Figure 5. Note that the similarity

threshold ρ_{th} of TPCE is empirically set to 0.9, meaning that TPCE will skip the candidates with a similarity smaller than 0.9, which explains the low detection rates of TPCE. A smaller ρ_{th} may increase both the detection rate and the run time. For example, setting $\rho_{th} = 0$ will lead to full search equivalence. We do not consider smaller ρ_{th} but use the parameters in [11] if not specify. As showed in Figure 5, comparing with ZNCC-based algorithms, the detection rates of PDE are less than 0.38 for Leuven, indicating that SAD is not robust to illumination changes. Our method yields the same or very close detection rates as the full search ZNCC-based algorithms (FFT and HFFT) for Leuven, Ubc and Graffiti, and performs better for Boat on sizes smaller than 100×100 indicating its robust to small in-plane rotation. However, our method obtains a lower detection rate than FFTs for Bikes on sizes larger than 50×50 because images with deep blur lose many textures and DAISY_LBPs are not good at discriminating local texture-less structures. The detection rate of our ROI extraction model for Boat is highest on sizes larger than 32×32 because the model is designed to be robust to small in-plane rotation ($[-15^\circ, 15^\circ]$) while TM is not. For speed-ups over FFTs, our method obtains the highest speed-ups for our model can eliminate 97.78% of the candidates. HFFT obtains the second highest speed-ups except for Ubc on the size 32×32 . The average computation elimination for TPCE and PDE are 90.18%, 73.40%, respectively, which explains that TPCE is faster than PDE. The average speed-ups of our method over FFTs, TPCE, PDE and HFFT are 30.90, 6.21, 24.08, 2.97 times, respectively.



FIGURE 4 Images from MIT database (the first row) and images from OX database (the second row)

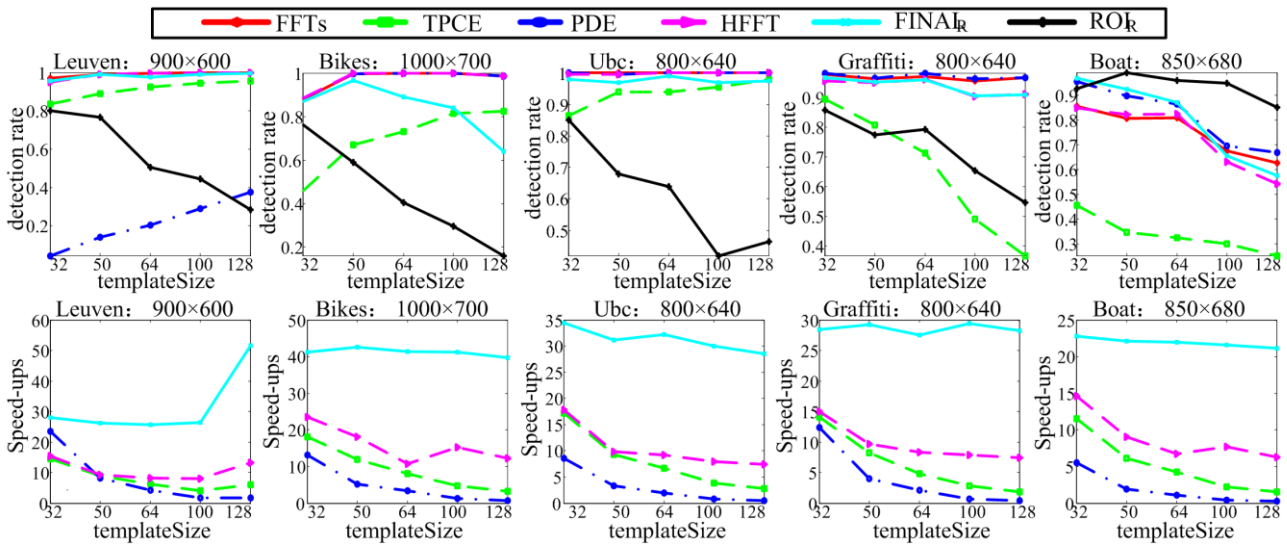


FIGURE 5 Detection rates and speed-ups over FFTs in run time for images with illumination changes (Leuven), blur (Bikes), JPEG compression (Ubc), small scale changes (Graffiti) and small in-plane rotation (Boat)

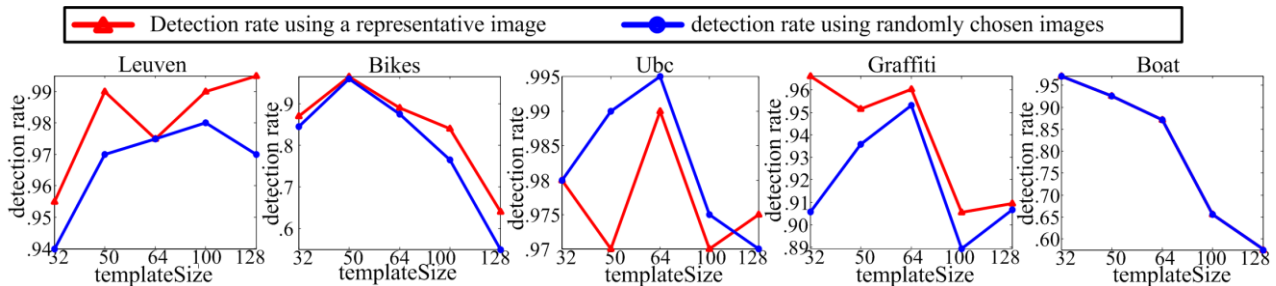


FIGURE 6 Comparison of the detection rate using a representative image and that using randomly chosen images to estimate the pattern distribution of the background

4.3 EXPERIMENTS WITH RANDOMLY CHOSEN IMAGES AS BACKGROUND

In the former experiments, a representative image is utilized to evaluate the pattern distribution of the background. However, there are cases that only a template is available. In this case, we can randomly choose a set of images to evaluate the pattern distribution of the background. In this experiment, 40 images mentioned in Section 4.1 are used to evaluate $p_{c,k}$, $k=0,1,\dots,65535$. We repeat the experiments in Section 4.2. The differences of detection rates are illustrated in Figure 6. Let " $p_{FINAL_{R1}}$ " and " $p_{FINAL_{R2}}$ " denote the detection rates in Section 4.2 and section 4.3 of the proposed method, respectively. As we can see, $p_{FINAL_{R2}}$ are exactly the same as $p_{FINAL_{R1}}$ for Boat, and very close to $p_{FINAL_{R1}}$ for Leuven, Ubc, Graffiti and Bikes. Let $|x|$ denote the absolute value of x . The maximums of $|p_{FINAL_{R2}} - p_{FINAL_{R1}}|$ are 0.025, 0.090, 0.020, 0.060, 0.000 for Leuven, Bikes, Ubc, Graffiti and Boat, respectively. In all, using randomly chosen images to estimate the pattern distribution of the background does not have obvious influences on the performance of our method. Using the pre-computation of $p_{c,k}$, $k=0,1,\dots,65535$, the training phase only needs to compute p_k , which will further reduce the training time to less than 0.20 seconds for template size 128×128 .

5 Conclusions

This paper proposes a fast top-down visual attention

method to downsize the search space of TM. A texture pattern namely DAISY_LBP is first introduced, which is efficient to compute and robust to noise and local distortions. The pattern is used in the top-down visual attention model, and the pattern weight describes how well a certain pattern is correlated to the specific template. Using patterns as the stimuli and the pattern weights as the strength of stimuli, the top-down control is realized by setting the pattern weights learned off-line. Experiments show that our method obtains 30.90, 6.21, 24.08, 2.97 times speed-ups over FFTs, TPCE, PDE and HFFT, respectively with little or no loss in performance.


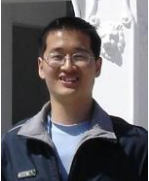


Our current method relies on a single ROI. In future work, several ROIs can be extracted to further improve the detection rate. The number of ROIs and the size of ROIs should make a compromise for efficiency. We will investigate the effects of these two terms. Efforts will also be given to the integration of colour and texture features into the algorithm for performance improvement.

Acknowledgments

This work is supported by National Natural Science Foundation of China (no. 61005028, no. 61175032 and No. 61101222).

References

- [1] Lewis J P 1995 Fast Normalized Cross-Correlation *Vision Interface* **10**(1) 120-3
- [2] Tang F, Tao H 2007 Fast Multi-Scale Template Matching Using Binary Features *IEEE Workshop on Applications of Computer Vision* 2007 36-41
- [3] Stefano L D, Mattoccia S, Tombari F 2005 *Pattern Recognition Letters* **26**(14) 2129-34
- [4] Montrucchio B, Quaglia D 2005 *IEEE Trans. on Circuits and Systems for Video Technology* **15**(2) 210-20
- [5] Ouyang W, Tombari F, Mattoccia S., Di Stefano L, Cham W 2012 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **34**(1) 127-43
- [6] Schweitzer H, Deng R, Anderson R A 2011 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **33**(3) 459-70
- [7] Pele O, Werman M 2008 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **30**(8) 1427-43
- [8] Tombari F, Ouyang W, Stefano L D, Cham W 2011 *Pattern Recognition Letters* **32**(15) 2119-27
- [9] Mattoccia S, Tombari F, Stefano L D 2008 *IEEE Trans. on Image Processing* **17**(4) 528-38
- [10] Mattoccia S, Tombari F, Stefano L D 2011 *Pattern Recognition Letters* **32**(5) 694-700
- [11] Mahmood A, Khan S 2012 *IEEE Trans. on Image Processing* **21**(4) 2099-108
- [12] Borji A, Itti L 2012 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **35**(1) 185-207
- [13] Carrasco M 2011 *Vision Research* **51**(13) 1484-525
- [14] Toet A 2011 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **33**(11) 2131-46
- [15] Duncan K, Sarkar S 2012 *IET Computer Vision* **6**(6) 514-23
- [16] Frintrop S, Backer G, Rome E 2005 Goal-Directed Search with a Top-Down Modulated Computational Attention System *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition* Berlin Heidelberg: Springer 117-24
- [17] Navalpakkam V, Itti L 2006 An Integrated Model of Top-Down and Bottom-Up Attention for Optimizing Detection Speed *IEEE Conference on Computer Vision and Pattern Recognition* 2006 **2** 2049-56
- [18] Fang Y, Lin W, Lau C, Lee B 2011 A Visual Attention model Combining Top-Down and Bottom-Up Mechanisms for Salient Object Detection *IEEE International Conference on Acoustics, Speech and Signal Processing* 2011 1293-6
- [19] Chang K, Liu T, Chen H, Lai S 2011 Fusing Generic Objectness and Visual Saliency for Salient Object Detection *IEEE International Conference on Computer Vision* 2011 914-21
- [20] Ma L, Cheng J, Liu J, Wang J, Lu H 2010 Visual Attention Model Based Object Tracking *Advances in Multimedia Information Processing* Berlin Heidelberg: Springer 483-93
- [21] Viola P, Jones M 2004 *International Journal of Computer Vision* **57**(2) 137-54
- [22] William P, Saul T, William V, Brian F 1992 *Numerical Recipes. The Art of Scientific Computing* Cambridge: Cambridge University Press
- [23] Ojala T, Pietikainen M, Maenpaa T 2002 *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**(7) 971-87
- [24] Liu L, Zhao L, Long Y, Kuang G, Fieguth P 2012 *Image and Vision Computing* **30**(2) 86-99
- [25] Pietikainen M, Hadid A, Zhao G Y, Ahonen T 2011 *Computer Vision Using Local Binary Patterns*, London: Springer
- [26] Winder S, Hua G 2009 Brown M Picking the Best Daisy, *IEEE Conference on Computer Vision and Pattern Recognition* 2009 178-85
- [27] Kanan C, Tong M H, Zhang L, Cottrell G W *Visual Cognition* **17**(6) 979-1003

Authors	
	<p>Yiping Shen</p> <p>Current position, grades: PhD student of University of Chinese Academy of Sciences University studies: University of Science and Technology of China (2005-2009).</p>
	<p>Shuxiao Li</p> <p>Current position, grades: Associate Professor, Institute of Automation, Chinese Academy of Sciences University studies: Xi'an Jiao Tong University (1999-2003), PH.D. on Pattern Recognition and Intelligent System (2008, Institute of Automation, Chinese Academy of Sciences). Scientific interest: computer science, machine vision, object recognition</p>
	<p>Chengfei Zhu</p> <p>Current position, grades: Assistant Professor, Institute of Automation, Chinese Academy of Sciences University studies: University of Science and Technology of China (2000-2004), PhD on Pattern Recognition and Intelligent System (2010, Institute of Automation, Chinese Academy of Sciences) Scientific interest: computer science, object recognition</p>
	<p>Hongxing Chang</p> <p>Current position, grades: Professor, Institute of Automation, Chinese Academy of Sciences, Dean of the Integral Information Research Center University studies: Beijing University of Aeronautics and Astronautics (1982-1986), Master (1991) Scientific interest: Integral information processing and intelligent system, computer vision</p>

Step semantics and action refinement in event structures

Weidong Tang ^{1, 2, 3}, **Jinzhao Wu** ^{1, 2, 3*}, **Meiling Liu** ^{1, 4}

¹*School of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China*

²*Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, China*

³*Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis, Nanning 530006, China*

⁴*Science Computing and Intelligent Information Processing of GuangXi higher education key laboratory, Nanning 530023, China*

Received 5 May 2014, www.tsi.lv

Abstract

An event structure acts as a denotational semantic model of concurrent systems. Action refinement is an essential operation in the design of concurrent systems. However, there exists an important problem about preserving equivalence under action refinement. If two processes are equivalent with each other, we hope that they still can preserve equivalence after action refinement. In linear time equivalence and branching time equivalence spectrum, step equivalences, which include step trace equivalence and step bisimulation equivalence are not preserved under action refinement [17]. In this paper, we define a class of concurrent processes with specific properties and put forward the concept of clustered action transition, which ensures that step equivalences are able to preserve under action refinement.

Keywords: event structure, action refinement, concurrency, step equivalence, clustered equivalence

1 Introduction

In order to model concurrent systems, we hope to have formal method for hierarchical structure. Action refinement is the core operation of the hierarchical method, which interprets an action in higher abstract layer with a process in lower layer, hence reduces the level of abstraction and eventually reaches its implementation layer. In the development course from top to bottom of concurrent system, we must first build models, which depict the system with description language of top layer; subsequently, according to these descriptions, we complete its implementation. This course often requires equivalence notion to verify the correctness of implementation of system. More concretely, assuming that P represents the descriptions of system and Q represents its implementation, if P is equivalent with Q (expressed as $P \approx Q$), then this shows that Q is correct. In development, the description P of a system can be refined layer-by-layer; accordingly, its implementation Q can be converted from framework into code or electronic components. Only the description and its implementation at all levels are required to maintain equivalence so as to ensure correctness of its implementation. This leads to an important question what kind of equivalence is maintained under action refinement, that is, if two concurrent systems are equivalent with each other, we hope that they still can preserve equivalence after action refinement.

Vogler [31, 33] first raised the basic thought of preserving equivalence under action refinement. Czaja, Van Glabbeek and Goltz [34] demonstrated that if interleaving bisimulation equivalence doesn't produce choice operations or action self-concurrences after actions are refined then it can preserve equivalence under action refinement, but interleaving trace equivalence still cannot preserve. Goltz and Wehrheim [30] proved that history preserving bisimulation is consistent with global causal dependencies, but they did not further discuss about the problem how to preserve equivalence under action refinement, and did not discuss that there are other situations under environment of action independencies. At last, in paper [17] Van Glabbeek and Goltz summarized that research results of action refinement in recent ten years, gave a detailed explanation for preserving equivalence problem under action refinement, and proved that interleaving equivalence cannot preserve under action refinement in general, but did not discuss further. Moreover, no work further discusses preserving problem under action refinement of step trace equivalence and step bisimulation equivalence. In this paper, we define a class of concurrent processes with specific properties and put forward the concept of clustered action transition, which ensures that step equivalences are able to preserve under action refinement in the absence of constraints.

2 Event structures and action refinement

Assume that Act be a set of actions.

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

Definition 2.1 A event structure \mathcal{P} is a 5-tuple $(E, <, \#, \Delta, l)$, where

E is the set of events;

$\leq \subseteq E \times E$ is irreflexive partial relation, and satisfy the

rule of finite causes that $\forall e \in E: \{e1 \in E | e1 < e\}$ is finite; In addition, its inverse “ $<$ ” is expressed as “ $>$ ”;

$\# \subseteq E \times E$ is irreflexive and finite conflicting relation, and satisfy the rule of inheriting of conflict that $\forall e1, e2, e3 \in E: e1 < e2 \wedge e1 \# e3 \Rightarrow e2 \# e3$;

$\Delta \subseteq E \times E$ is irreflexive concurrent relation, altogether with $<$ and $\#$ to satisfy the principle of partition that $< \cup \# \cup \Delta = E \times E$,

$e1 \Delta e2 \Leftrightarrow \neg(e1 = e2 \vee e1 < e2 \vee e2 < e1 \vee e1 \# e2)$;

$l: E \rightarrow Act$ is a label function of actions.

In this paper, let \mathbb{S} stand for the set of all event structures.

Definition 2.2 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. A relation between \mathcal{P} and \mathcal{Q} is called isomorphic (expressed as $\mathcal{P} \cong \mathcal{Q}$) iff there exists an bijection between their sets and preserves corresponding relations with $<, \#, \Delta$ and same corresponding labels.

Unless specified, we do not discriminate isomorphic event structures.

The behaviour of event structure is depicted with configuration which is the set of events with specific properties. Configurations are considered as possible states of system. The following is its definition.

Definition 2.3 Let X be a subset of the set $E_{\mathcal{P}}$ of all events in event structure \mathcal{P} .

(1) X is left closed iff $\forall e1, e \in E: e \in X \wedge e1 < e \Rightarrow e1 \in X$;

(2) X is conflicted-free iff $\mathcal{P}|_X$ is conflicted-free;

(3) X is a configuration iff X is not only left closed but also conflicted-free.

Here, let $C(\mathcal{P})$ represent the set of all configurations in event structure \mathcal{P} .

A configuration $X (X \in C(\mathcal{P}))$ is called successfully terminated configuration iff $\forall e \in E: e \notin X \Rightarrow \exists e1 \in X: e1 \# e$.

The event structure is also often represented with graph, where \rightarrow, \dots stands for casual relation and immediately conflict relation in event structure respectively, inheriting conflict relation, which is not considered and independent relation is not explicitly expressed.

Example 2.1 The system $\mathcal{P} = (a|b) + (c; b; d)$, executing either a, b concurrently or c, b and d sequentially, can be described by the event structure with events $e1, e2, e3, e4, e5$ with $l(e1) = a, l(e3) = c, l(e2) = l(e4) = b, l(e5) = d$, where $e1 \Delta e2, e3 < e4 < e5$,

each of $e1, e2$ is in conflict with each of $e3, e4, e5$. This event structure is expressed as below.

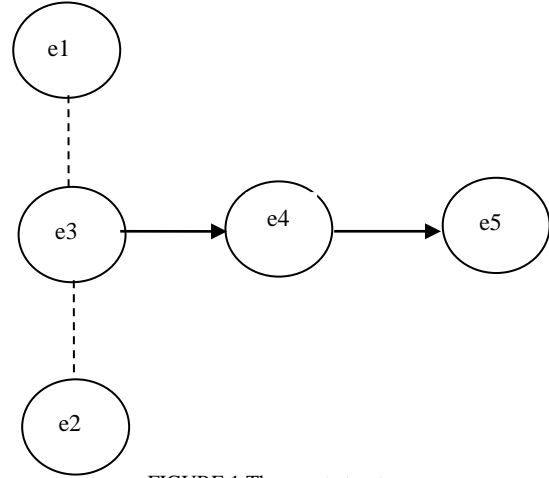


FIGURE 1 The event structure

Its configurations are $\emptyset, \{e1\}, \{e2\}, \{e1, e2\}, \{e3\}, \{e3, e4\}, \{e3, e4, e5\}$, where $\{e1, e2\}, \{e3, e4, e5\}$ are terminated configurations.

Basic thought of action refinement is: replace an action in higher layer with a process in lower layer, do it layer by layer, until get detailed design or implementation of system.

Definition 2.4 A function $ref: Act \rightarrow E - \{\emptyset\}$ is called a refinement function of event structure, iff $\forall a \in Act: ref(a)$ is not empty, finite and conflict-free. Let $\mathcal{P} \in \mathbb{S}$. $ref(\mathcal{P})$ is an event structure defined as follows:

$$E_{ref(\mathcal{P})} = \left\{ (e, e') \mid e \in E_{\mathcal{P}}, e' \in E_{ref(l_{\mathcal{P}}(e))} \right\}, \tag{1}$$

$$(e1, e1') <_{ref(\mathcal{P})} (e2, e2') \text{ iff } e1 <_{\mathcal{P}} e2 \text{ or } e1 = e2 \wedge e1' <_{ref(l_{\mathcal{P}}(e1))} e2', \tag{2}$$

$$(e1, e1') \#_{ref(\mathcal{P})} (e2, e2') \text{ iff } e1 \#_{\mathcal{P}} e2, \tag{3}$$

$$(e1, e1') \Delta_{ref(\mathcal{P})} (e2, e2') \text{ iff } e1 \Delta_{\mathcal{P}} e2 \text{ or } e1 = e2 \wedge e1' \Delta_{ref(l_{\mathcal{P}}(e1))} e2', \tag{4}$$

$$l_{ref(\mathcal{P})}(e, e') = l_{ref(l_{\mathcal{P}}(e))}(e'). \tag{5}$$

Example 2.2 Continue with Example 2.1. Assuming that $ref(b) = (b1; b2) + b3$, the event structure after action refinement is expressed as Figure 2.2.

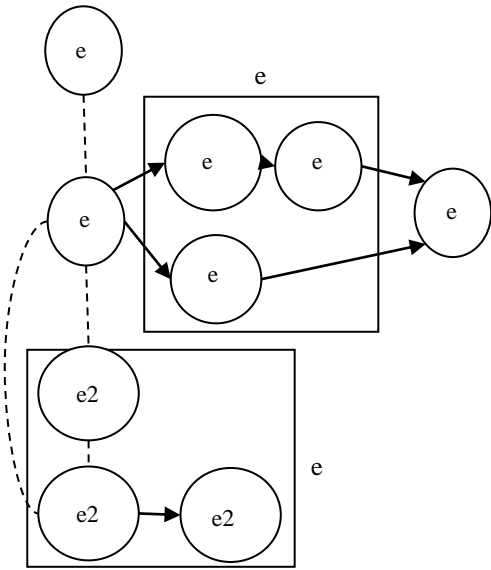


FIGURE 2 The event structure after action refinement

In the process of action refinement, each event e labelled by b is replaced by a disjoint copy, \mathcal{P}_e , of $ref(b)$, i.e. the event $e2$ is replaced by $(e22;e23)\#e21$ and the event $e4$ is replaced by $(e41;e42)\#e43$ [17]. The causality and conflict structure is inherited from \mathcal{P} : all events which were casually before e will be casually before all events of \mathcal{P}_e , every event, which casually followed e will casually follow all events of \mathcal{P}_e , and all events in conflict with e will be in conflict with all the events of all events which were casually before e will be casually before all events of \mathcal{P}_e .

3 Step equivalences

Step equivalences embody step trace equivalence and step bisimulation equivalence. To begin with, we give the definitions of single action transition and step action transition by comparison.

Definition 3.1 Let $\mathcal{P} \in \mathcal{S}$. A transition relation $X \xrightarrow{a} \mathcal{P} X'$ is called single action transition iff $a \in Act, X, X' \in C(\mathcal{P}), X \subseteq X'$, and $\exists e \in E_{\mathcal{P}} : X' - X = e, l_{\mathcal{P}}(e) = a$.

Here, $X \xrightarrow{a} \mathcal{P} X'$ denotes that the state expressed by Configuration X turns into the one expressed by Configuration X' after performing single action a in event structure $\mathcal{P} \in \mathcal{S}$.

Definition 3.2 Let $\mathcal{P} \in \mathcal{S}$. A transition $X \xrightarrow{A} X'$ is called a step action transition iff $A \in N^{Act}$ i.e., A is multiple set in $Act, X, X' \in C(\mathcal{P}), X \subseteq X', X' - X = G$, to make $\forall d, e \in G : d \Delta_{X'} e$ and

$l_{\mathcal{P}}(G) = A$, where $l_{\mathcal{P}}(G) \in N^{Act}$ is given by $l_{\mathcal{P}}(G)(a) = \{e \in G | l_{\mathcal{P}}(e) = a\}$.

Here, $X \xrightarrow{A} X'$ means that in event structure \mathcal{P} , after independently executing all actions of set A , the state expressed by configuration X is changed into the one expressed by configuration X' .

According to the above-mentioned definitions, obviously there is following proposition.

Proposition 3.1 A single action transition is also a step action transition.

Proof is omitted.

Then, we define trace and interleaving trace equivalence, step trace and step trace equivalence by comparison.

Definition 3.3 Let $\mathcal{P} \in \mathcal{S}$. A word $\mathcal{W} = a_1 \dots a_n \in Act^*$ is called a trace of event structure \mathcal{P} iff $\exists X_0, \dots, X_n \in C(\mathcal{P}) : X_0 = \emptyset$ and $X_{i-1} \xrightarrow{a_i} X_i, i = 1, \dots, n$.

Here, $trs(\mathcal{P})$ represents the set of all traces in event structure \mathcal{P} .

Definition 3.4 Let $\mathcal{P}, \mathcal{Q} \in \mathcal{S}$. A relation between \mathcal{P} and \mathcal{Q} is called interleaving trace equivalence (expressed as $\mathcal{P} \approx_{it} \mathcal{Q}$) iff $trs(\mathcal{P}) = trs(\mathcal{Q})$.

Definition 3.5 Let $\mathcal{P} \in \mathcal{S}$. A sequence $\mathcal{W} = A_1 \dots A_n (A_i \in N^{Act} (i = 1, \dots, n))$ is called a step trace of event structure \mathcal{P} iff $\exists X_0, \dots, X_n \in C(\mathcal{P}) : X_0 = \emptyset$ and $X_{i-1} \xrightarrow{A_i} X_i, i = 1, \dots, n$ is a step action transition.

Here, $steptrs(\mathcal{P})$ represents the set of all step traces of event structure \mathcal{P} .

Definition 3.6 Let $\mathcal{P}, \mathcal{Q} \in \mathcal{S}$. A relation between \mathcal{P} and \mathcal{Q} is called step trace equivalence (expressed as $\mathcal{P} \approx_{st} \mathcal{Q}$) iff $steptrs(\mathcal{P}) = steptrs(\mathcal{Q})$.

Furthermore, we define interleaving bisimulation equivalence and step bisimulation equivalence by comparison.

Definition 3.7 Let $\mathcal{P}, \mathcal{Q} \in \mathcal{S}$. A relation $R \subseteq C(\mathcal{P}) \times C(\mathcal{Q})$ is called a interleaving bisimulation between \mathcal{P} and \mathcal{Q} iff $(\emptyset, \emptyset) \in R$ and if $(X, Y) \in R$ then

$$\begin{aligned} X \xrightarrow{a} \mathcal{P} X', a \in Act &\Rightarrow \exists Y' : \\ Y \xrightarrow{a} \mathcal{Q} Y' \wedge (X', Y') \in R, \\ Y \xrightarrow{a} \mathcal{Q} Y', a \in Act &\Rightarrow \exists X' : \\ X \xrightarrow{a} \mathcal{P} X' \wedge (X', Y') \in R. \end{aligned}$$

A relation between \mathcal{P} and \mathcal{Q} is called interleaving bisimulation equivalence (expressed as $\mathcal{P} \approx_{ib} \mathcal{Q}$) iff there exists a interleaving bisimulation between them.

Definition 3.8 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. A relation $R \subseteq C(\mathcal{P}) \times C(\mathcal{Q})$ is called a step bisimulation between \mathcal{P} and \mathcal{Q} iff $(\emptyset, \emptyset) \in R$ and if $(X, Y) \in R$ then

$$\begin{aligned} X &\xrightarrow{A} \mathcal{P} X', A \in N^{Act} \Rightarrow \exists Y' : \\ Y &\xrightarrow{A} \mathcal{Q} Y' \wedge (X', Y') \in R; \\ Y &\xrightarrow{A} \mathcal{Q} Y', A \in N^{Act} \Rightarrow \exists X' : \\ X &\xrightarrow{A} \mathcal{P} X' \wedge (X', Y') \in R. \end{aligned}$$

A relation between \mathcal{P} and \mathcal{Q} is called step bisimulation equivalence (expressed as $\mathcal{P} \approx_{sb} \mathcal{Q}$) iff there exists a step bisimulation between them.

According to the definitions of trace equivalence and bisimulation equivalence, obviously there are following two propositions.

Proposition 3.2 $\mathcal{P} \approx_{ib} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{it} \mathcal{Q}$.

Proposition 3.3 $\mathcal{P} \approx_{sb} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{st} \mathcal{Q}$.

The paper [17] has showed that step bisimulation equivalence and step trace equivalence cannot preserve under action refinement. The following proposition shows that if no independency exists in event structure then step equivalences (include step trace equivalence and step bisimulation equivalence) are able to preserve under action refinement.

Proposition 3.4 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$, let ref be a refinement function. If independency between events in an event structure is such that $\Delta_{\mathcal{P}} = \Delta_{\mathcal{Q}} = \emptyset$ then

- (1) $\mathcal{P} \approx_{st} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q})$.
- (2) $\mathcal{P} \approx_{sb} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{sb} \text{ref}(\mathcal{Q})$.

Proof (1) Because $\Delta_{\mathcal{P}} = \Delta_{\mathcal{Q}} = \emptyset$, any transition in event structures \mathcal{P} and \mathcal{Q} is single action transition. Also, by proposition 3.1, $\mathcal{P} \approx_{st} \mathcal{Q} \Leftrightarrow \mathcal{P} \approx_{it} \mathcal{Q}$. By proposition 3.5(1) [31], $\mathcal{P} \approx_{it} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{it} \text{ref}(\mathcal{Q})$, namely, $\text{trs}(\text{ref}(\mathcal{P})) = \text{trs}(\text{ref}(\mathcal{Q}))$. In \mathcal{P} and \mathcal{Q} , every action label is exactly the same and is refined in the same way; moreover, $\Delta_{\mathcal{P}} = \Delta_{\mathcal{Q}} = \emptyset$. Hence, there exist same multiple action sets where actions are just concurrent in $\text{ref}(\mathcal{P})$ and $\text{ref}(\mathcal{Q})$. So we arrive at the conclusion that $\text{steptrs}(\text{ref}(\mathcal{P})) = \text{steptrs}(\text{ref}(\mathcal{Q}))$, namely $\text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q})$.

Proof (2) Because $\Delta_{\mathcal{P}} = \Delta_{\mathcal{Q}} = \emptyset$, any transition in event structures \mathcal{P} and \mathcal{Q} is single action transition. Also, by proposition 3.1, $\mathcal{P} \approx_{sb} \mathcal{Q} \Leftrightarrow \mathcal{P} \approx_{ib} \mathcal{Q}$. By proposition 3.5(2) [31], $\mathcal{P} \approx_{ib} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{ib} \text{ref}(\mathcal{Q})$. In \mathcal{P} and \mathcal{Q} , every action label is exactly the same and is refined in the same way; moreover, $\Delta_{\mathcal{P}} = \Delta_{\mathcal{Q}} = \emptyset$. Hence, there exist same multiple action sets where actions are all concurrent and branching time properties of corresponding action are just the same in $\text{ref}(\mathcal{P})$ and

$\text{ref}(\mathcal{Q})$. So we arrive at the conclusion that $\text{ref}(\mathcal{P}) \approx_{ib} \text{ref}(\mathcal{Q}) \Rightarrow \text{ref}(\mathcal{P}) \approx_{sb} \text{ref}(\mathcal{Q})$.

4 Clustered action transition

We introduce new one class action transition, where A is multiple set in action set Act and all actions within A independently perform with each other. We call this multiple set A as a clustered action and call this class transitions as clustered action transitions that is in fact one kind of special step action transition. With clustered action transitions, we construct two new types of equivalence.

Definition 4.1 Let $\mathcal{P} \in \mathbb{S}$. A transition $X \xrightarrow{A} X'$ is called a clustered action transition iff $A \in N^{Act}$ (i.e., A is multiple set in Act), $X, X' \in C(\mathcal{P})$, $X \subseteq X'$, $X' - X = G$, the events in set G satisfy:

- (1) entire independency of causes:
 $\forall d, e \in G : (d \Delta_{X'} e) \wedge (\{e_1 \in E_{\mathcal{P}} \mid e_1 \Delta_{\mathcal{P}} e\} \cup \{e\} = \{e_2 \in E_{\mathcal{P}} \mid e_2 \Delta_{\mathcal{P}} d\} \cup \{d\} = G)$ and $l_{\mathcal{P}}(G) = A$;
- (2) same causality:
 $\forall e_1 \in E_{\mathcal{P}} \setminus G, \exists e_2 \in G : (e_1 < e_2 \Rightarrow \forall e_3 \in G : e_1 < e_3)$
 $\vee (e_1 > e_2 \Rightarrow \forall e_3 \in G : e_1 > e_3)$;
- (3) same conflict relation:
 $\forall e_1 \in E_{\mathcal{P}} \setminus G, \exists e_2 \in G : (e_1 \# e_2 \Rightarrow \forall e_3 \in G : e_1 \# e_3)$.

Here, $l_{\mathcal{P}}(G) \in N^{Act}$ results from $l_{\mathcal{P}}(G)(a) = |\{e \in G \mid l_{\mathcal{P}}(e) = a\}|$.

Clustered action transition $X \xrightarrow{A} X'$ represents that after independently executing all actions of set A , the state expressed by configuration X is changed into the one expressed by configuration X' in event structure \mathcal{P} .

Example 4.1 In a simple process $K = (a \parallel b); (c \parallel d) + f$, its actions names corresponds to events e_a, e_b, e_c, e_d and e_f respectively. Assume that event structure model \mathcal{P} of process K is \mathcal{P}_K , figure 4.1 describes all single action transitions and configurations of \mathcal{P}_K , whereas figure 4.2 describes its all clustered action transitions and configurations. Find from comparison of two figures that when arriving at configuration $\{e_a, e_b, e_c, e_d\}$, there exists 10 possible transitions in figure 4.1, however there exists only 2 clustered action transitions in figure 4.2.

This example shows that clustered action transitions can simplify the system expressed by single action transitions. If applying this thought to simplify systems, a lot of good results may acquire.

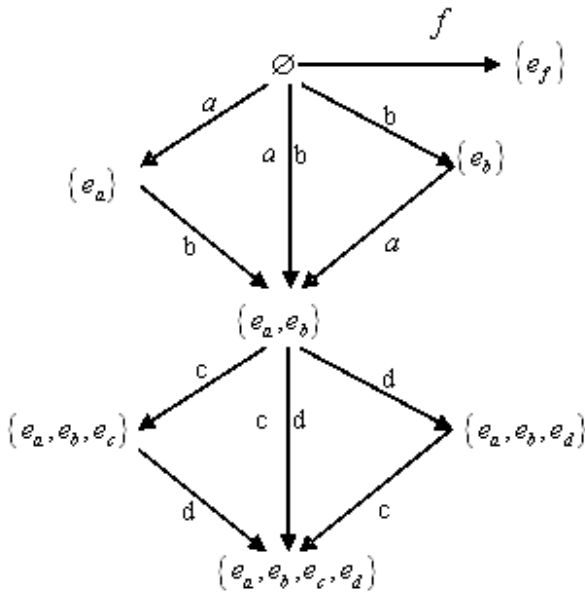


FIGURE 3 Single action transitions in a event structure

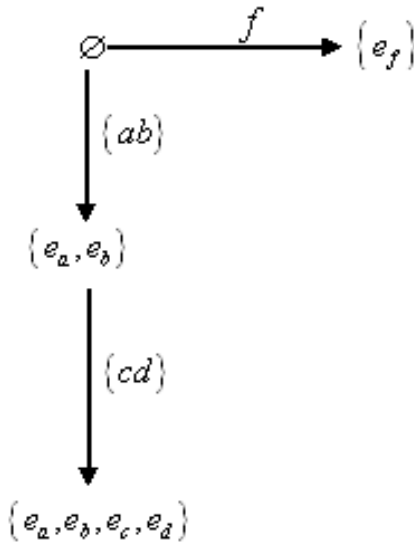


FIGURE 4 Clustered action transitions in a event structure

Definition 4.2 suppose that e is an event in event structure \mathcal{P} . Independency set of cause on e $\varphi(e)$, including itself of e , is defined as $\varphi(e) = \{e_1 \in E_{\mathcal{P}} \mid e_1 \Delta_{\mathcal{P}} e\} \cup \{e\}$.

In order to study the follow-up problems, we give a proposition in advance.

Proposition 4.1 Let $\mathcal{P} \in \mathbb{S}$. If all transitions in \mathcal{P} are clustered action transitions then

(1) $\forall d \in E_{\mathcal{P}} \Rightarrow \exists (X \xrightarrow{A} X') : \varphi(d) = G$, where $A \in N^{Act}$, $X, X' \in C(\mathcal{P}), X \subseteq X', X' - X = G$ and $l_{\mathcal{P}}(G) = A$;

(2) $\forall (X \xrightarrow{A} X') \Rightarrow \exists d \in E_{\mathcal{P}} : G = \varphi(d)$, where $A \in N^{Act}$, $X, X' \in C(\mathcal{P}), X \subseteq X', X' - X = G$ and $l_{\mathcal{P}}(G) = A$;

(3) $\bigcup_{e \in E_{\mathcal{P}}} \varphi(e) = E_{\mathcal{P}}$;

(4) $\forall e_1, e_2 \in E_{\mathcal{P}} : e_1 \neq e_2 \wedge \neg(e_1 \Delta e_2) \Rightarrow \varphi(e_1) \cap \varphi(e_2) = \emptyset$.

Proof Conclusions (1), (2) and (3) is very easy to reach, omitted here. Only give proof of (4). If $\varphi(e_1) \cap \varphi(e_2) \neq \emptyset$ then $\exists e \in \varphi(e_1) \cap \varphi(e_2)$. By definition 4.1, we obtain $\varphi(e) = \varphi(e_1) = \varphi(e_2)$, consequently obtain either $e_1 \Delta e_2$ or $e_1 = e_2$. This contradict with $e_1 \neq e_2 \wedge \neg(e_1 \Delta e_2)$. Accordingly, arrive at the conclusion that $\forall e_1, e_2 \in E_{\mathcal{P}} : e_1 \neq e_2 \wedge \neg(e_1 \Delta e_2) \Rightarrow \varphi(e_1) \cap \varphi(e_2) = \emptyset$.

The above proposition shows: If all transitions in an event structure are clustered action transitions then all independent actions involved in a clustered action transition are seen as a “big” action, its corresponding events can be thought of as a “big” event. Hence, not only no independency between events exists in this event structure but also independency sets of cause divide set of events into different parts, which induces an equivalence relation in set of events of this event structure.

Here, clustered action transition is a special kind of step transition. The fact that there exist entire independency of cause in a clustered action transition means that an action only belongs to a certain clustered action and is not shared with other clustered action. For example, in the process $(a;b) \parallel c$, action c may belong to two clustered actions i.e. $\{a,c\}$ and $\{b,c\}$, hence the event structure corresponding to this process cannot satisfy the requirement of entire independency of cause. Thus, the transitions here are not clustered action transitions but general step action transitions. This exactly is the reason why process $(a;b) \parallel c$ is step trace

equivalence with process $(a \parallel c); b + a; (b \parallel c)$, but step trace equivalence cannot hold under action refinement. On the contrary, independent events in clustered action transition possess same causal relation and same conflict relation, so every clustered action can be seen as a “big” action. These advantages ensure that equivalence based on clustered action transition can hold under action refinement. Accordingly, we introduce concept of clustered action transition equivalence.

Definition 4.3 Let $\mathcal{P} \in \mathbb{S}$. A sequence $\mathcal{W} = A_1 \cdots A_n$ ($A_i \in N^{Act}$ ($i = 1, \dots, n$)) is called a clustered trace of event structure \mathcal{P} iff $\exists X_0, \dots, X_n \in C(\mathcal{P}) : X_0 = \emptyset$ and $X_{i-1} \xrightarrow{A_i} X_i, i = 1, \dots, n$ is a clustered action transition.

Here, $\text{Clusteredtrs}(\mathcal{P})$ represents the set of all clustered traces of event structure \mathcal{P} .

Subsequently, we define clustered trace equivalence.

Definition 4.4 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$, all transitions in \mathcal{P} and \mathcal{Q} be clustered transitions. Let $\text{Clusteredtrs}(\mathcal{P})$ and $\text{Clusteredtrs}(\mathcal{Q})$ be the sets of all clustered traces of \mathcal{P} , \mathcal{Q} respectively. A relation between \mathcal{P} and \mathcal{Q} is called clustered trace equivalence (expressed as $\mathcal{P} \approx_{ct} \mathcal{Q}$) iff $\text{Clusteredtrs}(\mathcal{P}) = \text{Clusteredtrs}(\mathcal{Q})$.

The following propositions show that clustered trace equivalence is accordance with step trace equivalence in given conditions.

Proposition 4.2 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then $\mathcal{P} \approx_{ct} \mathcal{Q} \Leftrightarrow \mathcal{P} \approx_{st} \mathcal{Q}$.

Proof By definition 3.2 and definition 4.1, we draw the above conclusion at once.

Proposition 4.3 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then $\mathcal{P} \approx_{ct} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{it} \mathcal{Q}$.

Proof By definition 4.1 and proposition 4.1, $\forall A_i, A_j (i=1, \dots, n, j=1, \dots, n, i \neq j)$ in a clustered trace, then all actions in A_i are independent of those in A_j . Performing of each action in A_i does not interfere with those in A_j , vice versa. Consequently, $\text{Clusteredtrs}(\mathcal{P}) = \text{Clusteredtrs}(\mathcal{Q}) \Rightarrow \text{trs}(\mathcal{P}) = \text{trs}(\mathcal{Q})$, i.e. $\mathcal{P} \approx_{ct} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{it} \mathcal{Q}$.

Provide that $W = A_1 \cdots A_n$, where $A_i \in N^{Act}$ ($i=1, \dots, n$) be a clustered trace, and $|A_i|$ stand for the number of elements, then the clustered trace W corresponds to $|A_1|! \times |A_2|! \times \cdots \times |A_n|!$ general traces. In example 4.1, the clustered trace $\{a, b\} \{c, d\}$ corresponds to 4 (namely, $|\{a, b\}| \times |\{c, d\}| = 2 \times 2! = 4$) general traces $abcd, abdc, bacd, badc$.

After discussing clustered trace equivalence, we begin with another new class of equivalence and study whether they can maintain under action refinement or not. This class of equivalence is designated clustered bisimulation equivalence. The following provides for its definition.

Definition 4.5 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$, let all transitions in \mathcal{P} and \mathcal{Q} be clustered action transitions. A relation $R \subseteq C(\mathcal{P}) \times C(\mathcal{Q})$ is called a clustered bisimulation between \mathcal{P} and \mathcal{Q} iff $(\emptyset, \emptyset) \in R$ and if $(X, Y) \in R$ then

$$X \xrightarrow{A} \mathcal{P} X', A \in N^{Act} \Rightarrow \exists Y' : Y \xrightarrow{A} \mathcal{Q} Y' \wedge (X', Y') \in R;$$

$$Y \xrightarrow{A} \mathcal{Q} Y', A \in N^{Act} \Rightarrow$$

$$\exists X' : X \xrightarrow{A} \mathcal{P} X' \wedge (X', Y') \in R.$$

A relation between \mathcal{P} and \mathcal{Q} is called clustered bisimulation equivalence (expressed as $\mathcal{P} \approx_{cb} \mathcal{Q}$) iff there exists a clustered bisimulation between them.

Obviously, by definition 4.5, we come to a decision that $\mathcal{P} \approx_{cb} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{ct} \mathcal{Q}$.

Proposition 4.4 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then $\mathcal{P} \approx_{cb} \mathcal{Q} \Leftrightarrow \mathcal{P} \approx_{sb} \mathcal{Q}$.

Proof is omitted.

The above proposition shows that clustered bisimulation equivalence is consistent with step bisimulation equivalence under certain conditions.

Proposition 4.5 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then $\mathcal{P} \approx_{cb} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{ib} \mathcal{Q}$.

Procedure of proof is similar to that of proposition 4.3, omitted here.

5 Preserving of step equivalences

The paper [17] has demonstrated that step equivalence cannot preserve under action refinement in the general case. However, proposition 3.4 shows that step equivalence without concurrency can preserve under action refinement. This part will extend proposition 3.4 and show that in the presence of concurrency, step equivalence may preserve under action refinement in given conditions.

Proposition 5.1 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$, let ref be a refinement function. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then

$$(1) \mathcal{P} \approx_{ct} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q});$$

$$(2) \mathcal{P} \approx_{st} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q}).$$

Proof (1) By proposition 4.2, $\mathcal{P} \approx_{ct} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{st} \mathcal{Q}$.

By definition 4.1 and proposition 4.3, firstly, each clustered action transition corresponds to many single action transitions formed by interleaving performing of clustered actions; Secondly, when single action is refined, the corresponding clustered action is refined; Thirdly, Concurrent actions involved in per clustered action can be seen as one action, accordingly, their corresponding events also can be seen as one event. Therefore, by this treatment, there is not independency of cause in this event structure. Hence, we derive from proposition 3.4(1) that $\mathcal{P} \approx_{ct} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{st} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q})$.

(2) By proposition 4.2 and the above derivation process, we immediately reach a conclusion that $\mathcal{P} \approx_{st} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{st} \text{ref}(\mathcal{Q})$.

This proposition shows that if all transitions in an event structure are clustered action transitions then step trace equivalence can hold under action refinement.

Subsequently, we discuss the relationship between clustered bisimulation equivalence and step bisimulation equivalence under action refinement, as well as how to preserve step bisimulation equivalence under action refinement.

Proposition 5.2 Let $\mathcal{P}, \mathcal{Q} \in \mathbb{S}$, let ref be a refinement function. If all transitions in event structures \mathcal{P} and \mathcal{Q} are clustered action transitions then

$$(1) \mathcal{P} \approx_{\text{cb}} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{\text{sb}} \text{ref}(\mathcal{Q});$$

$$(2) \mathcal{P} \approx_{\text{sb}} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{\text{sb}} \text{ref}(\mathcal{Q}).$$

Proof (1) By proposition 4.4, $\mathcal{P} \approx_{\text{ct}} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{\text{st}} \mathcal{Q}$.

By definition 4.1 and proposition 4.5, firstly, each clustered action transition corresponds to many single action transitions formed by interleaving performing of clustered actions; Secondly, when single action is refined, the corresponding clustered action is refined; Thirdly, Concurrent actions involved in per clustered action can be seen as one action, accordingly, their corresponding events also can be seen as one event. Therefore, by this treatment, there are not independency of cause in this event structure. Hence, we derive from proposition 3.4(2) that $\mathcal{P} \approx_{\text{cb}} \mathcal{Q} \Rightarrow \mathcal{P} \approx_{\text{sb}} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{\text{sb}} \text{ref}(\mathcal{Q})$.

(2) By proposition 4.4 and the above derivation process, we immediately reach a conclusion that $\mathcal{P} \approx_{\text{sb}} \mathcal{Q} \Rightarrow \text{ref}(\mathcal{P}) \approx_{\text{sb}} \text{ref}(\mathcal{Q})$.

This proposition shows that if all transitions in an event structure are clustered action transitions then step bisimulation equivalence can hold under action refinement.

To sum up, if all transitions in event structures are clustered action transitions then two kinds of step equivalences, including step trace equivalence and step bisimulation equivalence, can hold under action refinement. Proposition 5.1 and proposition 5.2 extend proposition 3.4.

References

- [1] Wu J 2001 Action refinement in timed LOTOS *Proc. Of ASCM'01, World Scientific Publ.* 183-92
- [2] Aceto L, Hennessy M C B 1994 Adding action refinement to a finite process algebra *Information and Computation* **115** 179-247
- [3] Fecher H, Majster-Cederbaum M, Wu J 2002 Action refinement for probabilistic processes with true concurrency models *Lecture Notes in Computer Science* **2399** 77-94
- [4] Boudol G 1989 Atomic actions. *Bull. Eur. Ass. Theoret. Comput. Sci.* **38** 136-44
- [5] Busi N, van Glabbeek R J, Gorrieri R 1994 Axiomatizing ST bisimulation equivalence *Proceedings of the IFIP TC2/WG2.1/WG2.2/WG2.3 Working Conference on Programming Concepts, Methods and Calculi* 169-88
- [6] Castellano L, De Michelis G, Pomello L 1987 Concurrency vs. interleaving: An instructive example *Bull. Eur. Ass. Theoret. Comput. Sci.* **31** 12-5
- [7] Majster-Cederbaum M, Wu J 2003 Towards action refinement for true concurrent real time *Acta Informatica* **39** 1-47
- [8] Clarke E M, Grumberg O, Minea M, Peled D 1999 State space reduction using partial order techniques *STTT* **2**(3) 279-87
- [9] Darondeau P, Degano P 1993 Refinement of actions in event structures and casual trees *Theoretical Computer Science* **118** 21-48
- [10] Darondeau P, Degano P 1989 Casual trees *Automata, Languages and Programming, Lecture Notes in Computer Science* **372** 234-48
- [11] Jiang J, Wu J 2005 Symmetry and autobisimulation *Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and technologies. IEEE Computer Society Press* 866-70
- [12] Jiang J, Wu J, Yan W 2005 Structural reductions in process algebra languages *Proceedings of the 11th Joint International Computer Conference. World Scientific Publishing Co.* 596-600
- [13] Degano P, Gorrieri R 1995 A causal operational semantics of action refinement *Information and Computation* **122** 97-119
- [14] van Glabbeek R J, Goltz U 1989 Equivalence notions for concurrent systems and refinement of actions *Mathematical Foundations of Computer Science, Lecture Notes in Computer Science* **379** 237-48
- [15] van Glabbeek R J 1990 The linear time-branching time spectrum *CONCUR '90, Lecture Notes in Computer Science* **458** 297-8
- [16] van Glabbeek R J, Goltz U 1990 *A deadlock-sensitive congruence for action refinement* Institut fuer Informatik, Technische Universitaet Munchen:SFB-Bericht 342/23/90 A
- [17] van Glabbeek R J, Goltz U 2001 Refinement of actions and equivalence notions for concurrent systems *Acta Informatica* **37**(4/5) 229-327

6 Results and Discussion

The paper has demonstrated that (1) In event structures without independency between events, step trace equivalence and step bisimulation equivalence can preserve under action refinement; (2) In event structures, if all transitions are clustered action transitions, then with clustered trace equivalence between event structures, we can reach that step trace equivalence can preserve under action refinement; likewise, with clustered bisimulation equivalence between event structures, we can reach that step bisimulation equivalence can preserve under action refinement.

Therefore, we find a class of concurrent processes with specific properties, which enable step equivalence to preserve under action refinement in the absence of constraint.

Our next work is to introduce the thought of clustered action transition into model checking so as to deal with states explosion problem in the process of system verification.

Acknowledgements

This Work is supported by Grants No. HCIC201306 of Guangxi HCIC lab Open Fund, the National Natural Science Foundation of China under Grant No. 11371003, the Natural Science Foundation of Guangxi under Grant No. 2011GXNSFA018154 and No. 2012GXNSFGA060003, the Science and Technology Foundation of Guangxi under Grant No. 10169-1, the Scientific Research Project No. 201012MS274 from Guangxi Education Department., and Science Computing and Intelligent Information Processing of GuangXi higher education key laboratory No. GXSCIP201201.

- [18] Fecher H, Majster-Cederbaum M, Wu J 2002 Refinement of actions in a real-time process algebra with a true concurrency model *Electronic Notes in Theoretical Computer Science* **70**(3) 620-40
- [19] Gorrieri R, Rensink A 2001 *Action Refinement* //Bergstra J A, Ponse A and Smolka S A, editors Handbook of Process Algebra New York: Elsevier Science 1047-147
- [20] Huhn M 1996 Action refinement and property inheritance in systems of sequential agents *Concur'96, Lecture Notes in Computer Science* **1119** 639-54
- [21] Jategaonkar L, Meyer A R 1992 Testing equivalences for Petri nets with action refinement *Concur'92, Lecture Notes in Computer Science* **630** 17-31
- [22] Majster-Cederbaum M, Wu J 2001 Action refinement for true concurrent real time *Proc. ICECCS'01, IEEE Computer Society Press* 58-68
- [23] Majster-Cederbaum M, Wu J, Yue H 2006 Refinement of actions for real-time concurrent systems with causal ambiguity *Acta Informatica* **42**(6/7) 389-418
- [24] Majster-Cederbaum M, Wu J 2003 Adding action refinement to stochastic true concurrency models *ICFEM'03 Lecture Notes in Computer Science* **2885** 226-45
- [25] Alur R, Brayton R K, Henzinger T A, Qadeer S, Rajamani S K 1997 Partial-order reduction in symbolic state-space exploration *CAV'97, Lecture Notes in Computer Science* **1254** 340-51
- [26] Sun X, Zhang W, Wu J 2004 Event-based operational semantics and a consistency result for real-time concurrent processes with action refinement *Journal of Computer Science and Technology* **19**(6) 828-40
- [27] Winskel G 1989 *An Introduction to Event structures* Berlin: Springer, LNCS **354** 364-97
- [28] Wu J 2000 Logic programming-taking advantage of symmetry *Proc. Of ASCM'00, World Scientific Publ.* 100-9
- [29] Wu J, Fecher H 2004 Symmetric structure in logic programming *Journal of Computer Science and Technology* **19**(6) 803-11
- [30] Goltz U, Wehrheim H 1996 Modelling causality by dependency of actions in branching time semantics *Information Processing Letters* **59**(4) 179-84
- [31] Vogler W 1991 *Bisimulation and action refinement* STACS'91, Lecture Note in Computer Science **480** 309-21
- [32] Tang W, Wu J *Interleaving Semantics and Action Refinement in Event Structures* To be published
- [33] Vogler W 1992 Modular construction and partial order semantics of Petri nets *Lecture Note in Computer Science* **625** 625-48
- [34] Czaja I, van Glabbeek R J, Goltz U 1992 Interleaving semantics and action refinement with atomic choice *Advances in Petri Nets, Lecture Notes in Computer Science* **609** 89-107

Authors	
	<p>Weidong Tang</p> <p>Current position, grades: Associate professor University studies: Computer Software and Theory Scientific interest: Symbolic computation, formal verification</p>
	<p>Jinzhao Wu</p> <p>Current position, grades: Professor, Ph.D. University studies: Computer Software and Theory Scientific interest: Symbolic computation, automated reasoning, formal methods</p>
	<p>Meiling Liu</p> <p>Current position, grades: Teacher, Lecturer University studies: Computer Software and Theory Scientific interest: Data Mining, formal verification</p>

A research into static traffic routing and resource optimization algorithm based on genetic and tabu search

Xianfeng Yang¹, Yan Wang^{2*}

¹*School of Information Engineering, Henan Institute of Science and Technology, Henan Xinxiang, China*

²*Center of Modern Education Technology Xinxiang Medical University*

Received 7 May 2014, www.tsi.lv

Abstract

In order to solve the issue of optical network's static traffic routing and resource optimization, this paper puts forward a hybrid genetic and tabu search virtual reconfiguration algorithm (HGTS-VRA) and designs the key elements. This algorithm could effectively integrate the large scale searching ability of genetic algorithm and the outstanding local searching ability of tabu search algorithm. The simulation comparison result and analysis result show that the HGTS-VRA put forward by this paper enjoys excellent advantages in the field of traffic routing and resource optimizing. In addition, it offers outstanding extendibility and robustness.

Keywords: Hybrid Genetic, Tabu Search, Static Traffic Routing, Resource Optimization

1 Introduction

The virtual reconfiguration of the SDH/OTN/DWDM three level networks is a NP-Complete issue, which needs a mathematical model to express it in terms of mathematical formula from the aspect of engineering optimization, so as to determine the target and limitations for optimization. For the solution for issues within small scale network, it's recommended to adopt integral linear programming or mixed integral linear programming. However, with expansion of network scale and increase of connection request, the linear programming model is hard to gain the optimized solution within polynomial time. For the solution to the issues in large network, we need to adopt heuristic algorithm to meet to limited time requirement. However, though heuristic algorithm is able to give a solution within given time, it cannot offer the most optimized solution. As a result, the choice of algorithm becomes extremely important.

This paper researches into the static traffic routing issue within SDH/OTN/DWDM three level networks. In other words, it researches into the routing and resource optimization of static traffic. It can be divided into two sub-issues: namely the routing issue and network resource optimization issue, which are interdependent with close relationship. In the aspect of routing issue, while selecting route for a large batch of traffic, we need to take network resource optimization into comprehensive consideration. Therefore, the network resource optimization is actually within the selection of routing issues. In the aspect of network resource optimization, we need to reconstruct the virtual topology gained from solution of routing issue for the sake of

researching further into how to minimize resource consumption.

2 The Basic Elements of Genetic Algorithm (GA)

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 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

*Corresponding author e-mail: wy@xxmu.edu.cn

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 phynotype 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.

(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]:

$$p_i = \frac{F_i}{\sum_{i=1}^M F_i}, (i = 1, 2, \dots, M) \quad (1)$$

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) 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.

(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 Hybrid Algorithm Based on Genetic Algorithm and Tabu Search

3.1 THE IDEA OF HYBRID ALGORITHM

Theoretically, it has been proven that GA can find the optimal solutions to the problems in a random way from the significance of probability; however, the practical applications have also demonstrated that some unsatisfactory problems will appear in GA applications. The main problems include: easy to produce the premature phenomena; bad in local optimization and inefficient in running, which, however, are difficult to erase from GA. On the other hand, some optimization algorithms such as gradient method, hill-climbing method, simulated annealing algorithm and Tabu Search have strong local search capacity. Obviously, to mix the ideas of these optimization algorithms in GA search process and form a hybrid algorithm can enhance the running efficiency and the solution quality. The practice has shown that the improved GA is much better than the simple GA.

The hybrid GA blending the ideas of local search algorithm in the standard GA has the following two main characteristics:

(1) Introduce local search. Conduct local search based on the corresponding phenotypes to the individuals in the group and find the locally optimal solution to every individual in the current environment so as to improve the overall performance of the group;

(2) Add transcoding operation. Change the locally optimal solutions deduced from the local search process into new individuals through encoding to have a new group with better performance as the basis for the next-generation genetic evolution.

The basic constituent principles of the hybrid GA

The hybrid GA formed by the standard GA and other optimization algorithms shall abide by the following three principles:

(a) Adopt as much encoding of the original algorithm as possible;

(b) Use the advantages of the original algorithm;

(c) Improve the genetic operators. Design the genetic operators that can adapt to new encoding way and integrate the related inspirations to the problems in the operators; therefore, the hybrid GA cannot only preserve

the global optimization of GA, but it can also improve its running efficiency.

On the basis of the above-mentioned three principles, mix the standard GA and Tabu Search into a hybrid genetic algorithm. In the following passage, I will briefly introduce Tabu Search first.

3.2 TABU SEARCH (TS)

The basic idea of TS is: give a current solution (initial solution) and a neighbourhood and determine several candidate solutions in the neighbourhood of the current solution; if the corresponding target value to the optimal candidate solutions is better than the current optimal solution state, then ignore its tabu, replace it with the current solution and the optimal solution state; add the corresponding object into the tabu list (which is used to record the tabu of the candidate solutions) and change the tenure of the object in the tabu list; if no such candidate solutions exist, choose the untabued optimal state from the candidate solutions as the new current solutions, ignore their advantages and disadvantages with the current solutions; add the corresponding objects into the tabu list and change the tenure of every object in the tabu list; repeat the above search process till it satisfies the stopping criterion. As a simulation of human's thinking, Tabu Search can accept some solutions which are not so good through the tabu (which mean memory sometimes) of the locally optimal solutions so as to avoid local search [7, 8].

Since Tabu Search simulate the "memory" function in human's intelligence, it can memorize some solutions which have been checked lately and make them to be the tabu of the next solution by setting a flexible memory structure; thus, it can effectively avoid the circuitous search, improve the search capacity of the algorithm in the solution space and enhance the optimization efficiency and performance by despising some criteria to remit some tabued best states so as to realize the global optimization [9].

The specific steps of Tabu Search are classified as follows [10]:

(1) Randomly choose some solutions to put into the tabu search collection $T(s)$ according to a certain proportion;

(2) Take a solution s_i from the tabu search collection and set the tabu list and the optimal state as null;

(3) Produce several neighbourhood solutions s_i from the current solution, from which to determine several candidate solutions;

(4) Decide whether the candidate solutions have met the aspiration criterion. If they have, replace s_i with the optimal state S_{max} , which meets the criterion as the new current solution, namely, $s_i = S_{max}$; put the corresponding tabu objects to S_{max} in the tabu list; release the objects which have satisfied the tabu length from the tabu list;

replace “the optimal” state with S_{max} and turn to Step (6); otherwise turn to Step (5);

(5) Decide the taboo attributes of the corresponding objects to the candidate solutions and choose the optimal state of the untabued objects in the candidate solution collection as the new current solution; put the corresponding objects in the tabu list and release those, which have satisfied the tabu length;

(6) If it meets the requirements of Tabu Search on s_i , finish the search on s_i and turn to Step (7); otherwise, return to Step (3);

(7) If every solution in the tabu search collection T has been conducted Tabu Search, finish this process; otherwise return to Step (2).

In Step (1), the selection proportion can be among 0-100%. When choosing 0, ignore the Tabu Search. The higher the selection proportion, the better the performance improvement of GA solutions; however, at the cost of calculation time, when the selection proportion reaches a certain degree, the proportion increase will not improve the performance of the solutions [11]. The selection proportion can be adjusted according to the practical requirements before running the procedures. From the above, it can be seen that the focal function (to realize local search), the tabu objects, the tabu list and the aspiration criterion are key to Tabu Search [12]. Among them, the focal function adopts the idea of local search and it is used to realize neighbourhood search; the setting of tabu list and tabu objects have demonstrated the characteristic that the algorithm avoids circuitous search and the aspiration criterion is not only an award of the excellent state, but also a relaxation of the tabu strategy.

It is found that TS is faster than GA in the search speed; however, it is also noticed that TS relies greatly on the initial solutions. A better initial solution can help TS to find better solution in the solution space while a bad initial solution will lower the convergence speed of TS and the solutions TS find are relatively bad [13]. Therefore, people will usually use a certain algorithm such as heuristic algorithm to get a satisfactory initial solution to improve the performance of TS. Another shortcoming of TS is that its search is single-single operation, namely there can be only one initial solution in the search and it can just move one solution to another solution in every generation without operation on several solutions (group) in every generation, just like GA [14].

4 Three Level Network Structure Model

The multi-level network defined by IEFT refers to an area that is controlled by the unified control panel and offers one or multiple exchanging ability, multiple data plane exchange and supports to traffic engineering. The concept of multi-level gives its focus to data plane and is normally classified into different groups in accordance with its data exchanging ability. The actual network topology researched in this paper consists of integrated

network nodes and multi-wavelength fibre optic link, as indicated in Fig. 1. In actual network, some nodes may have already integrated OXC & ODUK electronic cross matric devices as well as DXC and corresponding layer interfaces, such as O3; some nodes only integrated OXC & ODUK electronic cross matric devices and corresponding layer interfaces, such as O4; some nodes are only equipped with OXC devices for traffic transmission and exchange, such as O6. After the actual network is divided based on logic, each integrated node will be classified as the corresponding node in different layers of the network and form the SDH/OTN/DWDM three-layer network structure. Here, the classification is the classification of logic sense, while different nodes in different layers of the network still have close interrelationship, which will be indicated by interlayer link. In fact, interlayer link is an abstract presentation of interlayer interface devices. The interlayer link mentioned in this paper include DWDM-OTN interlayer link, DWDM-SDH interlayer link and OTN-SDH interlayer link

SDH layer business is undertaken by SDH layer logic link and OTN layer business is undertaken by OTN layer logic link, while SDH logic link can be constructed by DWDM layer fibre optical link in terms of wavelength granularity or that the STM-N bandwidth granularity is constructed by OTN layer logic link. For example, in Fig.1, the SDH layer logic link could provide 2 kinds of resource transmissions: the first transmission is provided by the by O1→O7, O7→O8 and O8→O5 fibre optical links from DWDM layer in terms of wavelength granularity; the second transmission is provided by D1→D5 logic link in OTN layer in terms of STM-N granularity, while the D1→D5 logic link is provided by optical path constructed by O1→O2, O2→O3, O3→O4 and O4→O5 logic links in OTN layer in terms of wavelength granularity. As a result, the R1→R5 link can be constructed by D1→D5 logic link in OTN layer, while the D1→D5 logic link in OTN layer can be constructed by O1→O5 optical link in DWDM layer. Therefore, there is a nesting relationship among the links in SDH, OTN and DWDM layer.

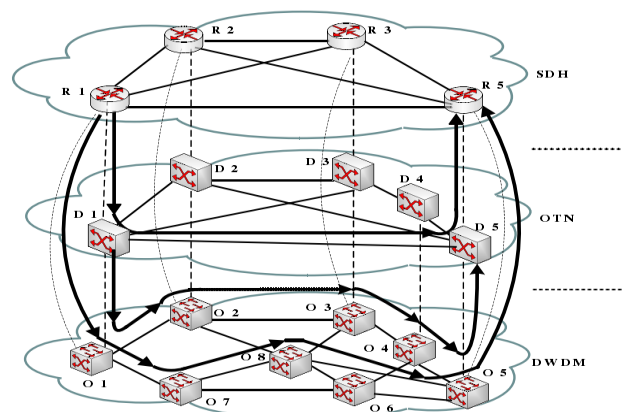


FIGURE 1 Multi-Layer Network Topology

5 Three Layer Optical Network

Under the circumstance of given network model parameters and a variety of limitations, setting the optimization target to request routing and resource configuration for traffic is the tyFigal integral programming issue. In order to verify the effectiveness of the algorithm put forward in this paper, we establish the static traffic routing issue optimization model named ILP. The accurate solution of this model could provide the theoretical upper performance limit of static traffic grooming algorithm.

5.1 MODEL PARAMETER

5.1.1 Symbol

1L, 2L, 3L: superscript or subscript, represent SDH layer, OTN layer and DWDM layer respectively.

- i, j: the two points of SDH layer virtual topology edge.
- u, v: the two points of OTN Layer topology edge
- m, n: the two points of DWDM layer optical link
- s, d: the traffic requests source node and host node of SDH layer and OTN layer

y: the band width granularity of traffic request. Here, both traffic in OTN layer and SDH layer are supposed to be fixed granularity:

OTN Layer Band Width Granularity $y \in \{ODU -1/2/3\}$, SDH layer bandwidth granularity $y \in \{STM -1/1/16\}$.

5.1.2 Known values

W: the maximum wavelength of each optical fibre

C^{2L-3L} : OTN Layer logic links is the maximum mapping link capacity that SDH layer can provide.

C^{1L-3L} , C^{1L-2L} : stand for OTN layer mapping link capacity and SDH layer mapping link capacity provided by DWDM layer wavelength links respectively.

$\Lambda_{2L}^{sd,y}$ and $\Lambda_{3L}^{sd,y}$: stand for the and width granularity traffic request number requested by OTN layer and SDH layer towards node (s, d).

R_j^{2L-3L} , R_j^{1L-3L} and R_j^{1L-2L} : standards for the Node J's interlayer link capacities from OTN layer to SDH layer, from DWDM layer to SDH layer and from DWDM layer to OTN layer.

T_j^{3L-2L} , T_j^{3L-1L} and T_j^{2L-1L} : stands for Node J's interlayer link capacity from SDH layer to OTN layer, from SDH layer to DWDM layer and from OTN layer to DWDM layer.

5.1.3 Variables

(1) DWDM layer topology variables

WOP_{mn}^{uv} : links from Node u to Node v within OTN layer and pass through Node M and Node N in DWDM layer mapping path.

WOP_{mn}^{ij} : links from Node i to Node j within SDH layer and pass through Node M and Node N in DWDM layer mapping path.

(2) OTN layer Topology Variables

$W_{uv,l}^{1L-2L}$ stands DWDM wavelength links mapped to DWDM layer by Node u's No. l link in OTN layer.

V_{uv}^{1L-2L} : stands for links from Node u to Node v in OTN layer and are mapped from DWDM layer to OTN layer

$SP_{uv}^{ij,y}$: The link group from Node i to Node j in SDH layer passes through Node u and Node v in OTN layer.

(3) SDH layer Topology Variables

$SP_{uv}^{ij,y}$: The link group from Node I to Node j in SDH layer and pass through Node u and Node v in OTN layer.

$W_{ij,k}^{1L-3L}$ stands for the DWDM wavelength that No. k link from Node I to Node j in SDH layer is mapped to DWDM layer.

V_{ij}^{1L-3L} and $V_{ij,y}^{2L-3L}$ stand for the links that locate in Node i in SDH layer and are mapped from DWDM layer (in terms of wavelength granularity) and OTN layer (in terms of y granularity) to SDH layer.

(4) Traffic Request Variables

$O_{sd} = 0$: $O_{sd} = 1$ indicates that, in Node (s, d) in OTN layer, the request of traffic Y of No. t granularity is accepted, otherwise, $O_{sd} = 0$

$O_{uv} = 0$: $O_{uv} = 1$ indicates that, in Node (s, d) in OTN layer, the request of traffic Y of No. t granularity is accepted, otherwise $O_{sd} = 0$

$S_{sd} = 0$: $S_{sd} = 1$ indicates that, in Node (s, d) in SDH layer, the request of traffic Y of No. t granularity is accepted, otherwise $S_{sd} = 0$

$S_{uv} = 0$: $S_{uv} = 1$ indicates, in Node (s, d) in SDH layer, the request of Traffic Y of No. t granularity is undertaken by links between Node (I, j) in SDH layer, otherwise $S_{uv} = 0$

5.2 OPTIMIZATION TARGET

The optimization modes are a target function that takes minimizing the network resource consumption as its target:

Min:

$$\{[\sum_{i,j} (C^{1L-3L} \times V_{ij}^{1L-3L}) + \sum_{u,v} (C^{1L-2L} \times V_{uv}^{1L-2L}) + \sum_{i,j,y} (C^{2L-3L} \times V_{ij,y}^{2L-3L})]\} \cdot (2)$$

Formula 2 stands for the interlayer link bandwidth consumed by traffic. Each new OTN layer logic link or SDH link needs assigned capacity from interlayer links and lower layer network links in accordance with

granularity. In addition, the consumed interlayer links are in terms of pair, including one downward interlayer link and one upward interlayer link. Therefore, the total consumed link width is the twice of total virtual topology logic links in OTN layer and SDH layer.

6 Adopt Three-layer Optical Transmission Network Static Topology in HGTS-VRA to Reconstruct

6.1 THE HYBRID ALGORITHM COMBINING GENETIC ALGORITHM AND TABU SEARCH

To genetic algorithm, once the individuals in the group are the same, no new genes can be introduced by selection and crossover algorithms and only mutation can transfer the group. When the mutation probability is small, the algorithm will linger on the old state for a long time; the search is inefficient and it is easy to converge in advance. Additionally, the selection operators make it more probable for the individuals with higher fitness value to survive; however, excessively-strong selection will over-attract the search process to the local minimum point, which makes it easy to converge in advance. The algorithm to define the target function by randomly taking the weights can find the approximate non-inferior solutions of various types, but the group will converge near several non-inferior solutions. Therefore, improvements should be made to the basic GA. Having flexible memory function and aspiration criterion and accepting inferior solutions in the search, Tabu Search has strong hill-climbing capacity and it can step out of the locally optimal solutions and turn to the other regions in the solution space in the search so as to increase the probability to obtain the better globally optimal solutions. Therefore, it is very necessary to introduce Tabu Search in GA; in the meanwhile, GA has the characteristic of global optimization and the final result does not depend on the selection of the initial value. On the other hand, Tabu Search searches along a line from a point and the quality and convergence speed of the final solution are closely related to the initial solution. Besides, Tabu Search can only have one initial solution in the search and it only transfers one solution to another solution in every generation without operating on the solution collection (group) like GA. Therefore, the hybrid strategy to combine genetic algorithm and Tabu Search can remedy each other's shortcomings and get a better optimization result.

This paper puts forward a Hybrid Genetic and Tabu Search Virtual Reconfiguration Algorithm (HGTS-VRA), which firstly adopts genetic algorithm for global search to determine to major areas with targeted individuals and secondly adopts tabu search algorithm in local search to improve individual quality. This algorithm adopts genetic algorithm in large-scale search. After the target area is narrowed to certain degree while all individuals' locations are fixed, the tabu search method is adopted for local search. In this way, we can significantly reduce the

times of calling tabu search algorithm and the amount of calculation. In addition, we can effectively integrate the genetic algorithm's advantage in large-scale search and tabu search algorithm's advantage in local search. The Fig.2 has presented the flow of this algorithm:

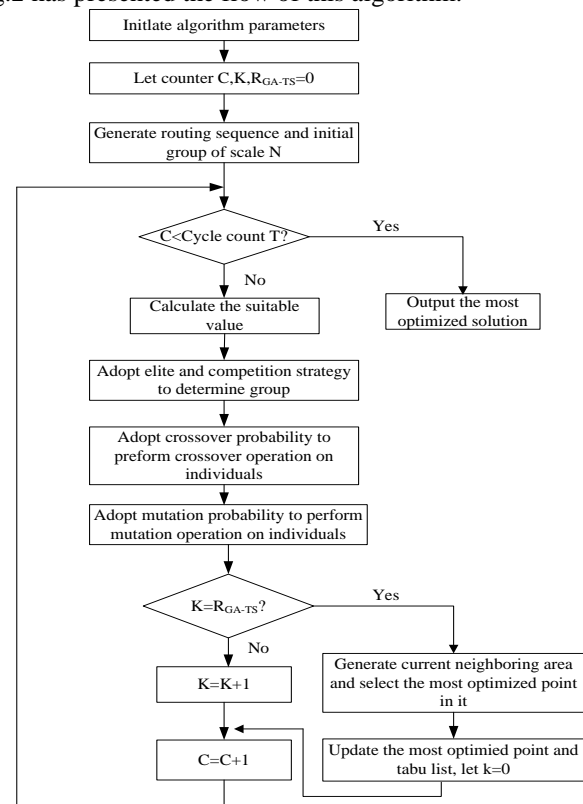


FIGURE 2 HGTS-VRA Algorithm Flow

6.2 ROUTING STEPS AND AUTHENTICATION SETTING

Static traffic routing is an issue global optimization and needs to request routing for batch static traffic of diversified bandwidth granularity. Such traffic request will be stored in terms of array. In regard to traffic request from OTN layer and SDH layer, this paper follows the sequence of requesting routing for traffic in OTN layer before routing SDH traffic and establishing logic links.

Step 1: For the routing of Traffic T_i (the No. i traffic request of OTN layer of SDH layer) in the virtual topology of current layer, we shall try to use rest bandwidth of logic links in current layer. In case of that, it fails in routing procedure, which we can turn to Step 2.

Step 2: Make comprehensive use of network resource in lower layer. The logic links that undertake $T_i(s, d, r)$ can adopt the mixed choice of current existed optic fibres and newly established optical fibre. In case that this routing fails, system will reject this traffic.

When a traffic request $T_i(s, d, r)$ is request routing in the three levels network, the weight value of Edge e ($W(e)$) in topology of DWDM layer, OTN layer and SDH layer shall be set as:

$$W(e) = \begin{cases} 1 & C(e) \geq r \\ \infty & C(e) < r \end{cases} \quad (3)$$

C(e) stands for the maximum rest band within e Links, C(e) guarantees that it will select the minimum tick for link among DWDM layer, OTN layer and SDH layer.

In regard to the traffic whose routing needs to take network resource in both upper layer and lower layer into comprehensive consideration, in case that they need to set different weight value to indicate the difference between link difference in different layers, the weight value of link i (W(i)) shall be set as:

$$W(e) = \begin{cases} \alpha \times C & C(i) \geq r \\ \infty & C(i) < r \end{cases} \quad (4)$$

C stands for the link width of interlayer links, while α stands for integrals larger than 1 to guarantee that W(i) >> 1

6.3 FITNESS FUNCTION

The fitness function is a reference to evaluate individuals in solved space's fitness in their environment. It has normally indicated or determined by target function, expense function or other method. We can set that, in the initial topology of SDH/OTN/DWDM, $G(V^1, E^1)$, $V^1 = \{v_i^1 | i = 1, 2, \dots, n\}$ and $E^1 = \{v_{ij}^1 | v_i^1, v_j^1 \in V^1\}$ stand for that node collection and optical link collection, while $e_{ij}^1 = 1$ or 0 indicates that $G(V^p, E^p)$ layer (2,3) virtual topology built by $G(V^1, E^1)$ respectively, $V^p \in V^1$ and E^p is the link grow of Layer p; set $L_{ij}^{pq} = \{l_{ij}^{pq} | v_i^p \in V^p, v_j^q \in V^q; p, q = 1, 2, 3; p \neq q\}$ stands for the link number array between layer p and layer q, while l_{ij}^{pq} stands for link number between Node i in layer p and Node j in Layer q. the fitness value is as below:

$$\varphi_f = I_{\max} - \ln \left[\sum_{p,q(p \neq q),i,j} l_{ij}^{pq} \right] \quad (5)$$

In this formula, I_{\max} is large enough figure to make $\varphi_f > 0$, while $\sum_{p,q(p \neq q),i,j} l_{ij}^{pq}$ stands for the interlayer resource usage situation.

6.4 CODING AND INITIAL GROUP GENERATING

One of the basic tasks of genetic algorithm is the coding of solution, which is also known as that the coding of one

solution is a chromosome, while the elements that build coding is known as gene. The purpose of coding is to optimize the solution presentation, so that it will be suitable for being calculated by genetic algorithm. The coding way of this paper: based on current network state of Topoi-1, in order to guarantee that traffic could select Topoi-1 shortest ways in Ti(s, d, r), it will randomly select one of the path to service as the path of this traffic and update Topoi-1 to be $Topo_i (i = 1, 2, \dots, m)$, while m stands for the total gene number in chromosome.

Generate the routing sequence for traffic in OTN layer and SDH layer respectively; select a rout for each traffic Ti(s, d, r) so as to form a chromosome; the mark for the path of each traffic is a gene from chromosome.

Repeat Step 2 N time to gain initial group of N.

6.5 GENETIC OPERATOR

Genetic operator is the main method to simulate group evolution. The genetic operation includes crossover operation, mutation operation and select operation. This paper introduces these three operations in details:

1) Crossover Operation

The crossover is the process in which two parent chromosomes produces new chromosome, while the crossover pattern of SDH layer traffic and OTN layer traffic is similar. Here, we cite OTN layer for example. Assuming that the capacity of new optical fibre is ODU-3 and the rest capacity of parent generation 2 A→B optical fibre 1 is ODU-2 while the rest capacities of other optical fibres are all ODU-2, the optical fibre A→E→D→C only undertakes Ti(A,C,ODU-2)and in Parent Generation 1, there are two optical links, namely optical fibre A→B and optical fibre B→C. In Parent Generation 2, there is a single fibre A→E→D→C. the crossover is the process of exchanging different chromosomes within the path selected by the same traffic based on current topology status. After performing crossover operation to the traffic, this business is undertaken by the new optical fibre A→E→D→C in Child generation 1. In Child generation 2, since the rest capability ODU-1 of optical fibre 1 between A→B is inferior to the bandwidth requested by Ti, the rest capacity of optical fibre 1 between A→B is not sufficient to undertake this traffic. However, the rest capacity of optical fibre 2 between A→B is larger than the band width requested by Ti and the rest capacity of optical fibre 1 between B→C is larger than the band width requested by Ti, this traffic will be undertaken by the optical fibre 2 between A→B and optical fibre 1 between B→C. since the optical fibre A→E→D→C in Child generation 2 does not undertake any traffic after crossover, the resource of this optical fibre will returned.

2) Mutation Operation

Mutation operation is to change a gene's path with steps as below:

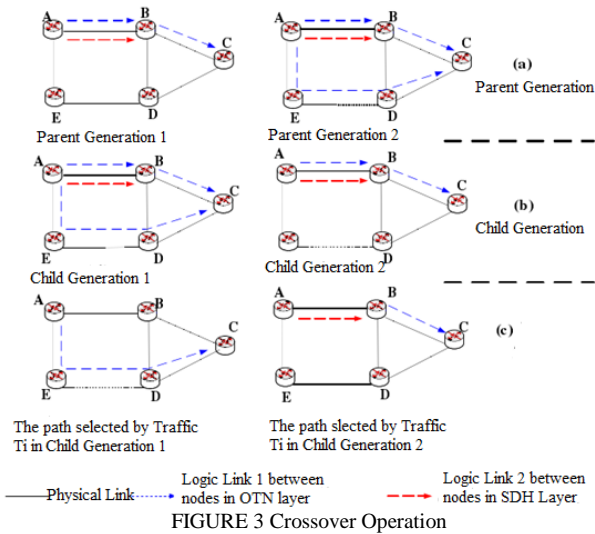


FIGURE 3 Crossover Operation

(1) Mutation based on traffic's tick in DWDM layer: since the logic links in SDH layer and OTN layer are undertaken by SDH layer and OTN layer, when a certain gene from chromosome take up less physical links in DWDM layer than this chromosome's average gene undertaking links, the significant mutation probability will be adopted, otherwise the small mutation probability will be adopted.

$$P_m^{(i,DWDM)} = \begin{cases} P_{mut} & (h_i \geq h_{avg}) \\ P_{mut} \times (h_{avg} - h_i) / (h_i + h_{avg}) & (h_i < h_{avg}) \end{cases} \quad (6)$$

P_{mut} is the initial value of mutation probability and h_i is the total physical links consumed by No. i gene in DWDM layer. $h_{avg} = \sum_{j=1}^m h_j / m$ is chromosome gene's average undertaking links in DWDM layer and m is total gene number in chromosome.

(2) Mutation based on the average logic link utilization ratio in SDH layer or OTN layer: when the utilization ratio of logic link in SDH layer or OTN layer virtual topology is not large than average utilization rate of logic links in virtual topology, the large mutation probability will be adopted so that the traffic in this link will re-choose path, otherwise, the small mutation probability will be adopted.

u_j is the utilization ratio of No. J logic link in virtual topology. $h_{avg} = \sum_{i=1}^q u_i / q$ is the average utilization rate of logic link, and q is the total logic link number.

3) Select Operation

Select operation is to select some outstanding individual from the current group. The standard of judging whether an individual is outstanding or not is their fitness value. This algorithm adopts the most popular tournament algorithm and saving best result algorithm. It can copy the individual solution of best fitness from current group to next group for comparison.

In this algorithm, the tournament algorithm steps are as below: randomly select 2 individuals from the group; compare the fitness of these two individuals; save the individual of highest fitness value to the next generation; repeat the above procedures until the individuals saved to next generation equals group number N.

7 Algorithm Simulation Result

We adopt NSFnet as the experimental network for comparing the performance of HGTS-VRA and LCBRF-StraGPP. The Fig4 below offers the performance comparison result between HGTS-VRA and LCBRF-StraGPP under the circumstance of different traffic distribution. LCBRF-StraGPP algorithm is an initial solution that takes LCBRF algorithm as tabu algorithm by using tabu node to achieving the most optimal algorithm.

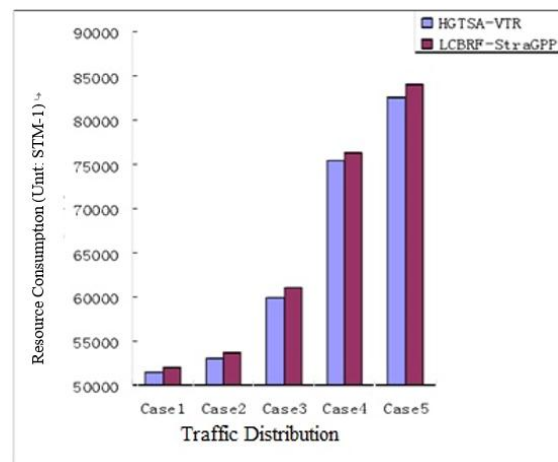


FIGURE 4 The Comparison of Network Resource Consumption between 2 Algorithms

Judged from Fig4, HGTS-VRA has better performance in the aspect of resource consumption compared with LCBRF-StraGPP. In addition, with the increase of traffic, this performance gap increases. HGTS-VRA's operation time is related to traffic volume, group quantity and frequency of calling tabu search. Since it adopt genetic algorithm and tabu mechanism at the same time, its operation time is around 3-5 times as that as LCBRF-StraGPP consumes. However, since it could the static traffic routing can be calculated offline, the decrease of resource consumption overtake the increase of operation time.

8 Conclusions

This paper puts forward a new optical fibre routing and resource optimization algorithm, which is known as the Hybrid Genetic and Tabu Search Virtual Reconfiguration Algorithm (HGTS-VRA). It can effectively integrate the large scale searching ability of genetic algorithm and the outstanding local searching ability of tabu search to help static traffic routing and resource optimization.

References

- [1] Goldberg 1989 *Genetic Algorithms in Search Optimization and Machine Learning* Addison-Wesley
- [2] Tsai C F, Tsai C W, Tseng C 2004 *A New Hybrid Heuristic Approach for Solving Large Traveling Salesman Problem* Oxford: Blackwell Scientific Publications **2004**(01)
- [3] DePuy G W, Whitehouse G E, Meta-Ra P S A Simple and Effective Approach for Solving the Traveling Salesman Problem *European Journal of Operational Research* **2009**(02)
- [4] Pelikan M, Mühlenbein H 2008 Marginal distributions in evolutionary algorithms *In Proceedings of the International Conference on Genetic Algorithms Brno Czech Republic: Technical University of Brno Publisher* **2008** 90-5
- [5] Rastegar R, Meybodi M R 2005 A study on the global convergence time complexity of estimation of distribution algorithms. *Lecture Notes in Computer Science* **2005** 441-50
- [6] Faigle U, Kem W 2002 Some convergence results for probabilistic tabu search. *ORSA Journal on Computing* **4**(1) 32-7
- [7] Glover F, Hanafi S 2009 Tabu search and finite convergence *Discrete Applied Mathematics* **119** 3-36
- [8] Salhi S 2012 Defining tabu list size and aspiration criterion within tabu search methods *Computer and Operation Research* **29** 67-86
- [9] Józefowska J, Waligóra Cx, Weglarz J 2007 Tabu list management methods for a discrete-continuous scheduling problem *European Journal of Operational Research* **137** 288-302
- [10] Perez J A M et. al. 2003 Variable neighbourhood tabu search and its application to the median cycle problem *European Journal of Operational Research* **151** 365-78
- [11] Salhi S 2007 Defining tabu list size and aspiration criterion within tabu search methods *Computers and Operations Research* **29** 67-86
- [12] Hageman J A et. al. 2010 Hybrid genetic algorithm-tabu search approach for optimising multilayer optical coatings *Analytica Chimica Acta* **490** 211-22
- [13] Young-Jae Jeon, Jae-Chul Kin 2011 Application of simulated annealing and tabu search for loss minimization in distribution systems *Electrical Power and Energy Systems* **26** 9-18
- [14] Zhang Defu, Deng Ansheng 2005 An effective hybrid algorithm for the problem of packing circles into a larger containing circle *Computers and Operations Research* **32**(8) 1941-51

Authors



Yang Xianfeng, born on January 12, 1978, Mengjin, Henan, China

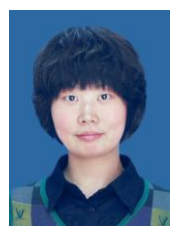
Current position, grades: lecturer

University studies: School of Information Engineering, Henan Institute of Science and Technology, Henan Xixiang, China

Scientific interest: pattern recognition, image processing, neural networks

Publications: Yang Xianfeng, Liuzhen, The Application of Association Rules Mining in Building Intelligent Transportation System, *Journal of Convergence Information Technology* 2012, Volume 7, Number 20.

Experience: Yang Xianfeng. He received his bachelor's degree in Computer Application Technology from Henan Normal University, Xixiang, Henan, China, in 2001, the master's degree in Computer Application Technology from China University of Petroleum, Dongying, China, in 2007. He is now a lecturer at School of Information Engineering, Henan Institute of Science and Technology, Xixiang, China. His current research interests include pattern recognition, image processing, neural networks, natural language processing.



Wang Yan, born on February 12, 1980, Taikang, Henan, China

Current position, grades: experimentalist

University studies: Centre of Modern Education Technology Xixiang Medical University, Henan Xixiang, China

Scientific interest: computer application technology ,image processing, computer-assisted two-and-three dimensional animation

Publications: Wang Yan, Wang Xiao, The Novel Analysis Model of Cloud Computing based on RFID Internet of Things, *SENSORS & TRANSDUCERS*, Vol.159-November 2013

Experience: Wang Yan. She received her bachelor's degree in Fine Arts from Henan University, Kaifeng, China, in 2002, the master degree in Fine Arts from Henan Normal University, Xixiang, China, in 2012. She is now a lecturer at Centre of Modern Education Technology in Xixiang Medical University. Her current research interests include computer application technology, modern education technology, medical image processing, computer-assisted two-and-three dimensional animation.

ABMP: adaptive bitmap protocol within TDMA for mobile underwater sensor networks

Xianmin Wei^{1*}, Hong Lu¹, Hong Feng²

¹School of computer engineering, Weifang University, 261061, Weifang, China

²School of Information Science and Engineering, Ocean University of China, 266100, Qingdao, China

Received 12 May 2014, www.tsi.lv

Abstract

Media access control (MAC) protocol design is one of hot topics in the research of underwater acoustic sensor networks (UWSN). The major challenge is the phenomenon of space-time uncertainty caused by the long delay in underwater signal propagation, where the occurrence of frame conflict is determined by not only two nodes' transmission time, but also by their locations. In this paper, the adaptive bitmap protocol-based (ABMP) within time division multiple access (TDMA) was proposed for UWSN, with specialized space-time uncertainty among mobile underwater nodes reducing channel idle time and improving energy efficiency and transmission efficiency. Finally, simulation experiments are conducted to present that the proposed protocol has better communication efficiency and energy efficiency, compared with other MAC protocols of Token-TDMA and T-lohi in terms of network traffic, end-to-end delay and energy efficiency.

Keywords: Underwater sensor networks, Adaptive bitmap protocol, Media access control, Time division multiple access

1 Introduction

Underwater sensor networks (UWSN) is the underwater extension of traditional terrestrial sensor network (WSN), which has important applications in a huge number of military and civilian areas, such as submarine detection, resource exploration, marine environmental monitoring and disaster warning, etc. Therefore, UWSN has been becoming one of the latest hot topics in recent years [1].

Compared with the conventional sensor network, most significant difference in UWSN is the communication method of acoustic signal as a carrier. Acoustic wave propagation characteristics in the sea has a huge difference from radio propagation in the air, which resulting in the unavailability of previous media access control (MAC) protocol in the UWSN.

1.1 CHARACTERISTICS OF UNDERWATER ACOUSTIC CHANNEL

Currently, there are numerous literatures on the propagation characteristics of the underwater acoustic channel study [2, 3], in general, the quality of underwater acoustic channel of communication is very poor, and its main characteristics are as follows:

(1) A large signal propagation delay. The average acoustic velocity in the sea is about 1,500 meters per second, which is several orders of magnitude lower than the propagation speed of the radio in the air, and acoustic wave propagation velocity changes with the

environmental impacts of ambient temperature, salinity, pressure and other factors.

(2) Low communication bandwidth. In the common communication distance, bandwidth is only a few dozen kb per second.

(3) High error-code rate. The reasons include the great background noise in the sea, and more commonly existing Doppler effects and multi-path effects. In underwater network, sound waves reflect while going through the seabed, sea surface and layered interface etc., to form multi-path effects (or multi-purpose effects); furthermore, the mobile communication nodes will cause the Doppler effect.

The remainder of the paper is organized as follows: Section 2 describes the related research on underwater MAC, while Section 3 presents detailed design of the protocol. Then, Section 4 describes the simulation results and protocol analysis, and finally Section 5 provides the conclusions.

2 Related research on underwater MAC

Currently, the underwater MAC protocols can be divided into three categories: competition-based protocols, non-competition TDMA protocol and other protocols.

2.1 COMPETITION-BASED PROTOCOLS

John Heidemann, first proposed the features of space-time uncertainty in hydro-acoustic network [5]: because of the high underwater acoustic channel propagation

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

delay, that whether the two nodes conflict or not, it is only depends on their transmission moment, but also on their geographic locations. And two targeted solutions were proposed: first, additional protection adding to the time slot Aloha protocol, secondly, T-lohi protocol [6] uses a competing channel with segment frame and uses back-off algorithm to avoid conflict.

Aloha-CA and Aloha-AN protocol [7] also uses the segment frame to compete channels, the main difference between them and T-lohi is that using virtual carrier sensing technology to avoid conflict without using back-off algorithm, the node maintains a neighbourhood state table, and on the basis of this table to adjust their transmission and reception. Aloha-CA split the data packet into shorter head and longer tail, and the head is used to compete for the channel, while Aloha-AN will not split the data packet using a separate channel segment frame to compete.

Slotted-FAMA protocol [8] is based on a handshake mechanism, which uses some slots working mode, and expands the length of negotiation slots to overcome the effects of space-time uncertainty, all of the control frames are ensured to be received within one time slot. Expansion length of negotiation slots to avoid conflicts leads to increase the idle channels and decrease network traffic, and therefore, channel utilization of slotted-FAMA does not exceed by 10%, while [9] has improved this issue.

2.2 UNDERWATER TDMA

Token-based TDMA protocol was proposed for dynamic constituted with autonomous underwater vehicles (AUV) [10], since it is difficult for the mobile network nodes to divide into a fixed communication time slot, nodes use clustering approach to manage and use tokens to get the transmission right.

TABLE 1 Comparison of three underwater TDMA protocols

ST-MAC	ECS	UW-FLASHR
Centralized	Distributed	Distributed
Time slot	No time slot	Timeslot
Short frame	Long frame	Long frame
Pre-assigned	Pre-assigned	Dynamically assigned

The basic idea of ST-MAC protocol [11] is assuming that underwater link delay is an integer multiple times than frame time (if the frame time is small enough); by assigning different transmission negotiation slots to avoid conflict. ECS [12] improved the ST-MAC, which allocating transmission moment for the nodes, abolishing the former assumptions and designed as distributed algorithm. UW-FLASHR [13] divides the underwater parallel transmission mode into three categories, and thus improves the efficiency of the conflict prediction. As for space-time uncertainty, better underwater solutions of these three TDMA protocols proposed are shown in the Table 1.

2.3 OTHER UNDERWATER PROTOCOLS

Using CDMA and FDMA technology faces many challenges in underwater sensor networks [14]. Wherein, TFO-MAC [15] is design for a single-hop network based on cluster multi-channel MAC protocol, while CDMA protocol based on tree topology was proposed in [16].

For AUV network a hybrid cluster protocol was proposed in [17], in which nodes in the same cluster using TDMA for communication, and between different clusters using CDMA to avoid interference. PLAN [18] is a hybrid protocol with CDMA and handshake mechanisms, by assigning different neighbours with CDMA orthogonal codes, multiple handshakes can be performed in parallel without conflict. A mixed protocol with Aloha and TDMA was presented in [19], the network switched based on load conditions: when the network load is light, using Aloha protocol of shorter delay, and when the network load is heavy, using TDMA protocol of non-conflict and heavy traffic.

2.4 RESEARCH TREND ON UNDERWATER MAC

As the underwater acoustic channel, communication environment is very harsh, and with characteristics of space-time uncertainty, underwater MAC protocol design generally has the following requirements to meet: high-energy efficiency (energy saving) [20], delays tolerance, robustness [22], reliability, flexibility (adaptation to the dynamic topology) [21]. To meet these requirements, the following means are mainly taken in current MAC protocol:

(1) Whenever possible, using a collision-avoidance scheme (e.g., TDMA) to save energy, in competition protocol, using short frame for competition or the reservation channel, instead of the manner of data packets direct competition, in order to ensure the transmission of data packets without conflict.

(2) For the characteristics of space-time uncertainty, use differences of different link delays to coordinate data transmission, and to improve the degree of parallelism in network transmission, and use clustering strategies and hierarchical topology management to deal with the dynamic changes of the network topology.

(3) Using a variety of techniques to improve transmission reliability. Such as CDMA and FDMA can solve hydro acoustic network problems of multipath effect and Doppler effect, using additional protection in the time slot scheduling, to reduce the probability of conflict in this line, and to ensure synchronization of only lightweight and local clock.

MAC layer is located at a lower level location in network protocol stack, which is directly affected by the physical layer properties, continuous deepening and understanding on the study of underwater acoustic physical channel and acoustic communication will bring new ideas for the MAC protocol design.

2.5 CHALLENGES FOR UNDERWATER MAC PROTOCOL DESIGN

Media access control (MAC) protocol determines the using mode of radio channel, in which has a significant impact on the efficiency of communication network. Due to the special nature of the underwater acoustic communication, many difficulties and challenges [4] facing underwater MAC protocol design is included as follows:

(1) Space-time uncertainty: the larger underwater delay difference leads to nodes' conflict, in which it is simultaneously determined by transmission time and their locations. However, the traditional MAC protocol considered to avoid conflicts only through the coordination of transmission time, which cannot be applied to the underwater environment.

(2) Energy restriction: UWSN are energy-constrained networks, energy saving has always been one of the core issues facing the protocol design. Underwater communication protocol must be established on the basis of energy saving, then it can be considered to improve the communication efficiency.

(3) Topology changing frequently: in addition to the nodes fixed on the sea bed, other nodes suspending or floating in the sea will move with the ocean currents, so UWSN is a dynamic network with topology changing frequently, network protocol design must adapt to this change.

Existing underwater MAC protocol considers less for mobile nodes, in fact, most of the suspending and floating nodes cannot be completely fixed, they will move within local small range with the ocean currents, waves, winds and other environmental factors. In this case, the propagation delay of the link changes dynamically. Therefore, this paper presents a bitmap-based dynamic TDMA protocol for this dynamic underwater sensor networks.

Because of the space-time uncertainty, frame segments from different sites sending simultaneously arrive at different time without any conflict, nodes complete negotiations and channel reservation by taking advantage of this particular phenomenon, and reduce the communication delay from end to end by dynamically adjusting TDMA cycle length. Simulation results show that the proposed protocol has better communication efficiency and energy efficiency.

3 Detailed design of the proposed protocol

This section details the design of the protocol, including the basic idea of the protocol, workflow, and performance optimization, and performance is analysed theoretically.

3.1 THE PROPOSED PROTOCOL PROCESS

Hydro acoustic network characteristic of space-time uncertainty is shown in Figure 1, in Figure 1(a), the nodes

A and C are sending a message to node B, A's transmission time is earlier than that of C, but the arrival time of A is later than that of C, this is due to the link delay of AB is much larger than that of BC, in which leading to "early sending and late arriving". In Figure 1(b), D and E simultaneously send messages to each other while they receive a message at the same time without conflict, this is because link delay between D and E is much greater than the time frame. This special phenomenon is unthinkable for traditional radio network with almost negligible link delay. We can take advantage of this particular phenomenon to perform a network protocol service in the hydro acoustic network, short frame collision probability is very low.

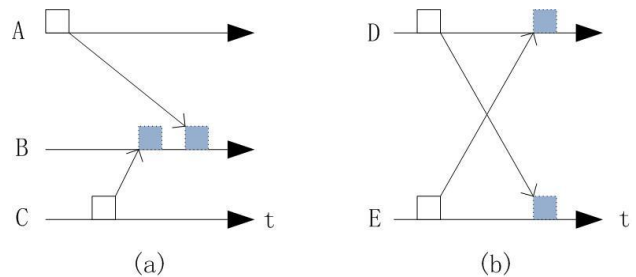


FIGURE 1 Example of space-time uncertainty

The basic process of the proposed ABMP within TDMA is shown in Figure 2. The protocol is composed of alternative negotiation time slot and TDMA transport period, and the transmission cycle is constituted of one or more negotiation slots, in which the length is dependent on the amount of data transmission required by the current nodes.

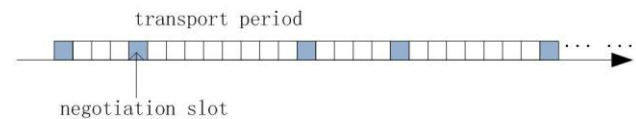


FIGURE 2 Protocol process of the proposed ABMP

Any frame should be transmitted at the beginning of the time slot, and guaranteed to be successfully received in the current time slot. Therefore, the length of the time slot is decided by the largest link propagation delay in the network, and the local nodes' movement under considerations. By definition, $Dis(i, j)$ is the space distance between the two points i and j , $Z(A)$ is the mobile node region of A , $S(A)$ is set of neighbouring nodes of A , T_{slot} is the time slot length in data transport period. Assuming that underwater acoustic signal propagation speed is v , T_{data} is the time to send a data packet (i.e., the frame time of data packet), then T_{slot} can be calculated by Equation (1). This definition is applicable to any underwater sensor networks with one-dimensional, two-dimensional and three-dimensional mobile nodes.

$$L_{max} = \max\{\max\{Dis(i, j) | i \in Z(A), j \in Z(B), \forall B \in S(A)\} | \forall A \in UWSN\}$$

$$T_{slot} = \frac{L_{max}}{v} + T_{data} \tag{1}$$

The commonly known partial mobile nodes are shown in Figure 3, in which neighbouring nodes A and B floating in the sea are anchored to the sea bed, so that their movement range is a circle, this kind of mobile node with small movable range is called restricted nodes. It shows that in Figure 3 the longest communication distance between A and B is Dis(C, D), while maximum of the maximal communication distance among all links in the network determines the length of T_{slot} .

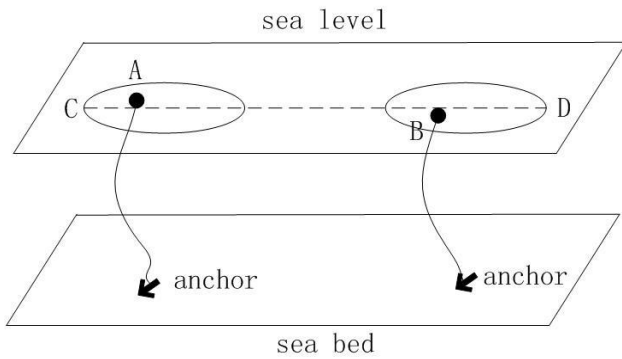


FIGURE 3 Calculation of the time slot length

During the time slot of consultation, each node which having data to transmit sends a short frame to reserve transmission order in the transport period. The short frame contains only three parameters: the source node ID, the destination node ID and a random number P, P is called the temporary priority, and which is calculated by the pseudo-random number generator as shown in Equation (2). The temporary priority is used to determine the node's transmission order in a data transport period.

$$P_{n+1} = aP_n + b(\text{mod } m). \tag{2}$$

Because of the existence of space-time uncertainty, nodes can receive the appointment short frame of other nodes with a great probability and without conflict; these nodes obtain the temporary priority of other nodes. In the data transfer phase, nodes can send data in the priority order of size, each node occupies the period of one slot time. Thus, the length of data transmission period is depended either on the number of nodes to be transmitted, or on the current network load.

Defined that T_{neg} is length of negotiation time slot, T_{short} is frame time required to send short frame of reservation ($T_{short} \ll T_{data}$), then T_{neg} is calculated by the following Equation (3).

$$T_{neg} = \frac{L_{max}}{v} + T_{short}. \tag{3}$$

3.2 IMPROVEMENT AND OPTIMIZATION OF NEGOTIATION TIME SLOT

In the stage of negotiation time slot, short frame sent by nodes will conflict in the both scenarios below:

Scenario 1: two or more nodes have the same temporary priority. As the proposed protocol in Equation (2) generates random number using linear identical-residue generator, as long as the parameters of a, b, and m are set sufficiently large, the occurrence probability of this situation is almost zero and can be almost ignored. When this case happens, the node ID can be transmitted as a second level priority parameter.

Scenario 2: short frame of two or more nodes conflict at the place of the receiving node, the node cannot receive the conflict frame at this time and cannot obtain priority information related to the node, and which is unable to complete an appointment and consultation. As shown in Figure 4, nodes A and B send information to C, assuming that at the current time the link propagation delay of AC and BC is D_{AC} and D_{BC} , respectively, frame time of the short frame is T_{short} , the short frame conflict condition at this time is shown in Equation (4).

$$|D_{AC} - D_{BC}| < T_{short} \tag{4}$$

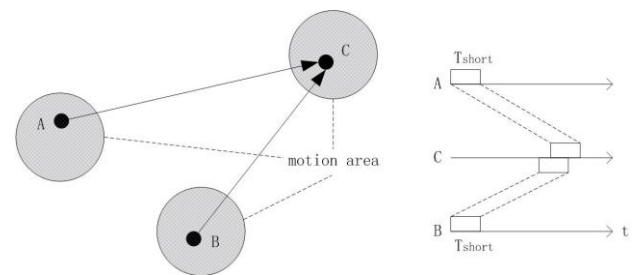


FIGURE 4 The conflict condition of short frame.

As the conflict of Scenario 2 occurs, the conflict nodes will lose the opportunity of channel reservation, and wait to resend short frame for negotiation in the next negotiation cycle. And the current transport period will waste one time slot, which is due to that other nodes without conflict are not unaware of this node's conflict, and will reserve one time slot for this node well in the current transport period according to its priority. The even worse situation is that, because in the ocean, the movement of the restricted mobile nodes has continuity and uncertainty, such a conflict state of short frame may continue for a long period.

Therefore, this subsection optimizes and improves the consultation slots to eliminate the conflict situation of the short frame. Before the beginning of the data transfer phase, nodes affected by conflict broadcast a warning message, then all nodes restart appointment. While reservation, the nodes with conflict run back-off algorithm to calculate their T_{wait} of back-off time, lastly before sending short frame the nodes prefer the awaiting time as T_{wait} shown in Equation (5). The re-appointment process is repeated until all the nodes in the network does not exist conflicts.

$$T_{wait}(i) = T_{short} \times \text{random}[0, \min(2^i, N_{max})]. \tag{5}$$

Equation (5) is a calculating formula of back-off time, wherein T_{short} is the transmission frame time of a short frame, integer N_{max} is the upper limitation value of the contention window size, n is the number of participating in the negotiation.

Thus, the reservation negotiation slots is expanded into reservation period, the dynamic reservation period can eliminate continuous frames conflict caused by underwater nodes' movement, and it ensures the fairness of channel reservation, i.e., the transmission order of nodes depends only on their temporary priority, and is not affected by the conflict.

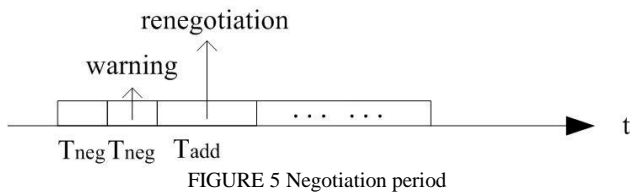


FIGURE 5 Negotiation period

Composition of negotiation cycle is shown in Figure 5, wherein T_{add} denotes the length of negotiation slots when required renegotiation, its value is calculated by Equation (6).

$$T_{add} = \frac{L_{max}}{v} + T_{short} \times N_{max} \quad (6)$$

3.3 NETWORK TRAFFIC ANALYSIS

Defined that T_n is the length of negotiation period, the length of the data packet is T_{data} , R is the ratio of T_{data} to T_{slot} , η is the channel utilization. Assuming that in the current data transfer phase, there are N nodes transmitting data, then η is calculated by Equation (7).

$$\eta = \frac{N \times T_{data}}{T_n + N \times T_{slot}} = \frac{1}{\frac{T_n}{N \times T_{data}} + \frac{1}{R}} \quad (7)$$

The above Formula presents that the channel utilization is related with parameters of R and T_n in the protocol. η decreases with the increase of T_n , and increases with the increase of R . When conflicts of scenario 1 and scenario 2 of short frame do not occur, T_n is a constant value. However, when the conflict occurs, because of the randomness and irregularity of nodes movement in the underwater environment, it is more difficult to predict and change the value of T_n .

In short, under the conditions allowed by the network, the packet length is increased as much as possible (i.e., increasing T_{data} and R) to increase network traffic and achieve greater channel utilization.

4 Simulation results and analysis

This section describes the simulation results of the proposed protocol compared to other protocols of T-lohi and Token-TDMA. Where T-lohi is a representative of underwater competitive MAC protocol, while most of the existing non-competitive protocols cannot be applied in a dynamic network, then token-TDMA protocol is the representative of non-competitive protocols.

Simulation results mainly investigate the normalized throughput, normalized end-to-end delay, and energy efficiency under different conditions of traffic load with the three protocols. The number of nodes needed in transmission within a fixed period of time is on behalf of the network load conditions, the network traffic is represented with the amount of information (number of bits) transmitted per time unit, the unit of end-to-end propagation delay is the number of slots (T_{slot}), energy efficiency is the ratio of the energy consumed to transmit a packet to total energy the network consumption.

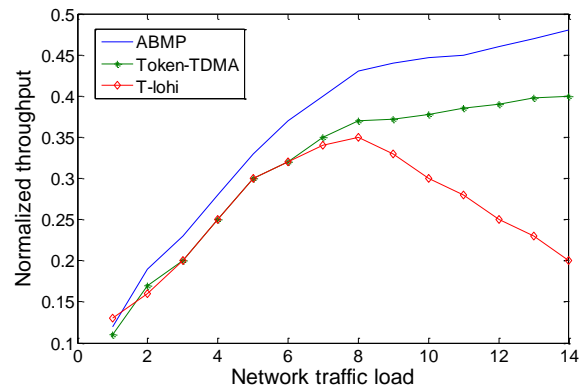


FIGURE 6 The network traffic when $R = 0.5$

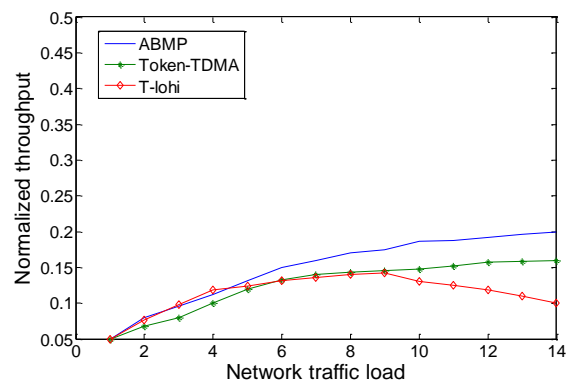


FIGURE 7 The network traffic when $R = 0.2$

Figures 6 and 7 show that normalized throughput as the function of network traffic load corresponding with different parameter values of R , the higher value of R , the larger network traffic, which verifies the flow analysis results in the above section. Results from Figure 6 and 7 present that the proposed protocol ABMP has the highest network traffic among the three MAC protocols. Under heavy load conditions, competitive-based protocols conflict more frequently, the performance of T-lohi protocol is the worst, traffic is minimum. For Token-

TDMA, when a node begins to transmit data till it has a token, so it must wait the first node in cluster head to assign a token, the waiting time increases the channel idle period, and reduces the channel utilization, so the flow is lower than protocols of ABMP.

Figure 8 shows the normalized end-to-end communication delay of the three protocols, when the network load is light, end-to-end delay of T-lohi is minimum, but it increases more greatly while the network load increases. Overall, protocol of ABMP has the smallest average end-to-end delay, due to the adaptation of dynamic transmission cycle strategy, avoiding the wait time of idle channel, when the network load grows, end-to-end delay only increase linearly.

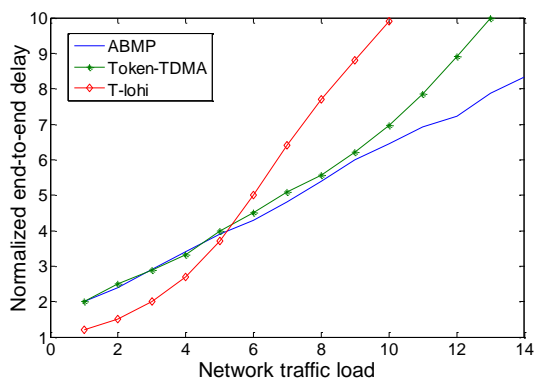


FIGURE 8 Normalized end-to-end delay

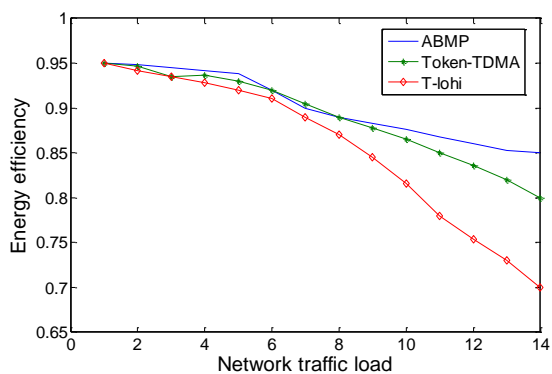


FIGURE 9 Energy efficiency

Figure 9 shows the energy efficiency of the three protocols, because the three protocols use short frame to compete or reserve channels, there does not exist conflicts in the data transfer phase, therefore which waste little energy, so its energy efficiency is higher and better than the other MAC protocols without using short frame reservation.

The proposed protocol of ABMP within TDMA has higher energy efficiency compared with these three protocols, which indicates that it uses fewer short frames; the efficiency of channel reservation is higher, because of

References

- [1] Li J, Gao H 2008 Survey on sensor network research *Journal of Computer Research and Development* 45(1) 45-1-1215
- [2] Guo Z, Luo H, Hong F, Yang M, Ni M 2010 Current progress and research issues in underwater sensor networks *Journal of Computer Research and Development* 47(2) 255-68

using the characteristics of space-time uncertainty, short frames in the protocol of ABMP conflict less.

5 Conclusions

In this paper, we propose the ABMP TDMA protocol, which is for suitable underwater network with restricted mobile node. The protocol uses hydro-acoustic network feature of space-time uncertainty, the node is able to use only one time slot to complete the channel reservation and consultation together, thereby to increase the network traffic and reduce the end-to-end communication delay. In the data transfer phase, the length of time slot can ensure that data is not affected by node movement and conflict-free communication. In addition, the proposed protocol is distributed algorithm, which does not rely on the dispatch from centre node. Simulation results show that, compared with MAC protocol in conventional underwater mobile network, ABMP has a smaller end-to-end delay, the higher energy efficiency and the higher network traffic.

Overall, the study of underwater acoustic network is still in its infancy, the MAC protocol design will be further deepen and complete with gradual understanding improvement on characteristics of underwater acoustic channel. Underwater MAC protocol design shows the following trends: emphasis on cross-layer design, integrated optimization for data transmission through multiple protocol layers of physical layer, data link layer and network layer to further improve network performance; due to the real complex underwater acoustic communication environment, computer simulations and laboratory simulations on underwater acoustic communication experiment have limitations, so the researchers pay more attention to experimental research and verification in real underwater environment, however, for expensive acoustic network test equipment's and the high cost of experiments, it is difficult to spread widely, which is now one of the bottlenecks in underwater sensor network research to be overcome.

Acknowledgments

This work is partly supported by National Natural Science Foundation of China (60933011, 61170258), Shandong Natural Science Foundation (ZR2011FL006), Shandong Science and Technology Development Plan (2011YD01044), Shandong Spark Program (2012XH06005), and Weifang municipal Science and Technology Development Plan (201301050).

- [3] Cui J, Kong J, Gerla M, Zhou S 2006 Challenges: Building scalable mobile underwater wireless sensor networks for aquatic applications *IEEE Network, Special Issue on Wireless Sensor Networking* 12-8
- [4] Heidemann J, Li Y, Syed A, Wills J, Wei Y 2006 Underwater sensor networking: Research challenges and potential applications

- In Proceedings of the IEEE Wireless Communications and Networking Conference* April Las Vegas, Nevad, USA
- [5] Syed A, Ye W, Krishnamachari B, Heidemann J 2007 Understanding spatio-temporal uncertainty in medium access with aloha protocols *In Proceedings of the Second ACM International Workshop on UnderWater Networks (WUWNet)*, ACM September Montreal, Quebec, Canada
- [6] Syed A A, Ye W, Heidemann J 2008 T-lohi: A new class of MAC protocols for underwater acoustic sensor networks *In IEEE INFOCOM*
- [7] Chirdchoo N, Soh W, Chua K 2007 Aloha-based MAC protocols with collision avoidance for underwater acoustic networks *Proc. IEEE InfoCom*, May
- [8] Molins M, Stojanovic M 2006 Slotted FAMA: A MAC protocol for underwater acoustic networks *In Proceedings of the IEEE OCEANS'06, Asia Conference*, May Singapore
- [9] Peleato B, Stojanovic M 2006 A mac protocol for ad-hoc underwater acoustic sensor networks *In WUWNet '06: Proceedings of the 1st ACM international workshop on Underwater networks*, New York, NY, USA, ACM Press 113-5
- [10] Li B, Jie H, Zhou S 2008 Further results on high-rate MIMO-OFDM underwater acoustic communications *Proc of IEEE OCEANS 2008, Piscataway, NJ*: IEEE 11-6
- [11] Zhong Z, Le S, Cui J 2010 An OFDM Based MAC protocol for underwater acoustic networks *Proceedings of the Fifth ACM International Workshop on Underwater Networks (WUWNet '10)*
- [12] Kim J, Lee J, Jang Y, Son K, Cho H 2009 A CDMA-Based MAC Protocol in Tree-Topology for Underwater Acoustic Sensor Networks *In proc. 2009 International Conference on Advanced Information Networking and Applications Workshops*. Bradford, United Kingdom, May 26- 29
- [13] Burrowes G E, Khan J Y 2009 Adaptive Token Polling MAC Protocol for Wireless Underwater Networks *Proceedings of the 4th international conference on Wireless pervasive computing (ISWPC'09)*. NJ, USA
- [14] Hong L, Hong F, Guo Z, Li Z 2011 ECS: Efficient Communication Scheduling for Underwater Sensor Networks *Sensors* 11(3) 2920-38
- [15] Yackoski J, Shen C 2008 UW-FLASHR: Achieving High Channel Utilization in a Time-Based Acoustic MAC Protocol *Proceedings of the third ACM international workshop on Underwater Networks (WUWNET'08)*
- [16] Salvá-Garau F, Stojanovic M 2003 Multi-Cluster Protocol for Ad Hoc Mobile Underwater Acoustic Networks *In Proc. IEEE Oceans* San Diego, CA, USA 91-8
- [17] Tan H, Seah W K G 2007 Distributed CDMA-based MAC Protocol for Underwater Sensor Networks *In proc. 32nd IEEE Conference on Local Computer Networks (LCN 2007)*, Dublin, Ireland, October 26-36
- [18] Kredó K B II, Mohapatra P 2007 A Hybrid Medium Access Control Protocol for Underwater Wireless Networks *In Proceedings of the Second Workshop on Underwater Networks*, Montreal, QC, Canada, 14 September 33-40
- [19] Hong L, Hong F, Guo Z, Guo Y 2011 HCR: A Hybrid MAC Protocol for Underwater Sensor Networks Using Channel Reservation *International Journal of Computers and Applications (IJCA)* 2(33) 154-9
- [20] Guo Z, Guo Y, Hong F, Jin Z, He Y, Feng Y, Liu Y 2010 Perpendicular Intersection: Locating Wireless Sensors with Mobile Beacon *IEEE Transactions on Vehicular Technology (TVT)* 59(7) 3501-09
- [21] Guo Z, Chen P, Feng Y, Jiang Y, Hong F 2010 ISDP: Interactive Software Development Platform for Household Appliances Testing Industry *IEEE Transactions on Instruments and Measurement*, 59(5) 1439-52
- [22] Guo Z, Li Z 2001 Lifetime Prolonging Algorithms for Underwater Sensor Networks *China Ocean Engineering* 20(2) 325-34

Authors	
	<p>Xianmin Wei, born on December 16, 1969 in Weifang, China</p> <p>Current position, grades: Associate Professor University studies: Shandong Science and Technology University Scientific interest: intelligent computing and intelligent sensor networks, swarm intelligent Publications: 20 papers Experience: He received the M. Sc. degree in computer applications from Shandong Science and Technology University (2005). He is currently an associate professor in school of computer engineering at Weifang University, China. He has published over 30 papers and 3 books in professional fields. Since 2011, he has been a member of IEEE-CS, ACM and CCF, respectively.</p>
	<p>Lu Hong, born on March 19, 1984 in Jining, China</p> <p>Current position, grades: lecturer, Ph.D. University studies: Ocean University of China Scientific interest: underwater protocols and algorithms in wireless sensor networks Publications: 10 papers Experience: He received the Ph.D. degree in School of Information Science and Engineering, Ocean University of China in 2011. He is currently a lecturer in school of computer engineering at Weifang University, China. He has published over 10 papers and obtained 3 patents in professional fields.</p>
	<p>Feng Hong, born on May 9, 1977 in Qingdao, China</p> <p>Current position, grades: Associate Professor University studies: Shanghai Jiaotong University Scientific interest: grid computing, wireless sensor networks and mobile delay-tolerant networks. Publications: 15 papers Experience: He received the Ph.D. degree in Shanghai Jiaotong University of China in 2009. He is currently an associate professor in School of Information Science and Engineering, Ocean University of China. He has published over 30 papers and obtained 3 patents in his professional fields.</p>

Design of grey PID controller for DC Servo Motor

Hong Li¹, ZongZhe Wang^{2*}

¹*School of Information Engineering, Xi'an Yang Normal University, Shaanxi, 712000. P. R. China*

²*Northwest Electric Power Construction Corporation, Shaanxi, 712000. P. R. China*

Received 1 May 2014, www.tsi.lv

Abstract

Concerning the uncertainties may be existed in DC motor servo system and the control quality that the external disturbance may influence the control algorithm of traditional PID, a kind of PID control algorithm based on grey prediction theory was proposed. With the grey theory's processing ability on the unknown information data, the algorithm established the grey model for the uncertainties, and real-time compensated the system's unmodelled feature and disturbance signal, thus improving the control precision. The simulation result can show that the proposed PID control algorithm based on grey prediction theory can effectively predict and compensate the unmodelled feature and disturbance signal in DC motor servo system and improve the control precision of the controller, thus providing the theoretical basis for the industrial application of PID control algorithm based on grey prediction theory.

Keywords: PID control, DC motor, Noise, Disturbance, Grey prediction

1 Introduction

With the feature of simple structure, small size, lightweight, reliable operation, easy maintenance, good servo performance, fast response speed and good stability, the DC motor is widely applied in the fields of servo system, factory automation and defence industry [1]. However, the measurement noise commonly existed in the industrial field, the structural disturbance caused by the abstract and simplification when establishing the model and the parameter disturbance caused by the parameter change in the actual work of the servo will seriously influence the working performance of the servo system [2]. The existence of above noise and disturbance makes the traditional PID control strategy fail to meet the control demand of DC motor servo system [3]. In order to solve the influence of above measurement noise and disturbance on the system control feature, the measurement noise and disturbance need to be predicted and compensated effectively. Grey prediction is a good choice to predict the above measurement noise and disturbance. It was firstly proposed by Professor Deng Julong from Huazhong University of Science and Technology [4]. Through the small amount of incomplete information, the grey differential prediction model was established to conduct the fuzzy and long description on the development law of the things. Introducing the grey prediction in servo control system can effectively predict the above measurement noise and disturbance, and conduct the corresponding compensation scheme according to the prediction result.

Concerning the uncertainties may be existed in DC motor servo system and the control quality that the

external disturbance may influence the control algorithm of traditional PID, the author proposed a kind of PID control algorithm combined the grey prediction theory and traditional PID control algorithm and applied in the DC motor servo system. With the grey theory's processing ability on the unknown information data, the algorithm established the grey model for the uncertainties, and real-time compensated the system's unmodelled feature and disturbance signal, thus improving the control precision.

The main structure of the paper is as follows: Section 2 established the mathematical modelling of DC motor; Section 3 described the theoretical basis for the grey prediction; the design of grey PID controller was described in Section 4; Section 5 was the numerical analysis and conclusion discussion.

2 Mathematical modelling of DC motor

In modern industry, DC motor is the implementation terminal with the most widely application in servo system. With the research object of brushless electromagnetic DC motor, the paper established the mathematical model [5]. The working principle of DC motor controlled by armature is as shown in Figure 1 [5].

According to Figure 1, the following equation can be obtained:

* *Corresponding author* e-mail: 110508@sina.com

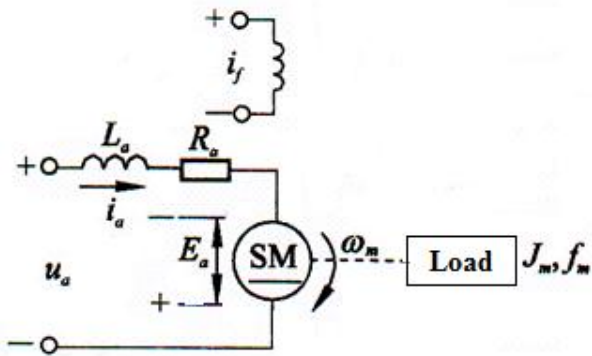


FIGURE 1 Working principle of DC motor

(1) Armature circuit voltage balance equation:

$$u_a(t) = L_a \frac{di_a(t)}{dt} + R_a i_a(t) + E_a \tag{1}$$

In the formula, E_a is the counter electromotive force of the armature. It is produced when rotating the electric digital. The size is proportional to the excitation flux and revolving speed, and the direction is opposite to the armature voltage $u_a(t)$.

(2) Electromagnetic torque equation:

$$M_m(t) = C_m i_a(t) \tag{2}$$

In the formula, C_m (N.m/A) is the moment coefficient of the motor. $M_m(t)$ (N.M) is the electromagnetic torque produced by the armature current.

(3) Torque balance equation of motor shaft:

$$J_m \frac{d\omega_m(t)}{dt} + f_m \omega_m(t) = M_m(t) - M_c(t) \tag{3}$$

In the formula, f_m (N. m/rad/s) is the viscous friction coefficient converted by the electromotor and load in the motor shaft. $J_m s$ (kg.m.s²) is the rotational inertia converted by the electromotor and load in the motor shaft.

Erasing the intermediate variables $i_a(t)$, E_a and $M_m(t)$ in (1)-(3), $\omega_m(t)$ can be obtained as the output quantity, and the DC motor differential equation with the input quantity of $u_a(t)$ is:

$$L_a J_m \frac{d^2 \omega_m(t)}{dt^2} + (L_a f_m + R_a J_m) \frac{d\omega_m(t)}{dt} + (R_a f_m + C_m C_e) \omega_m(t) = C_m u_a(t) - L_a \frac{dM_c(t)}{dt} - R_a M_c(t) \tag{4}$$

For the armature circuit inductance L_a is smaller in the engineer, it can be neglected; therefore, the above formula can be simplified as:

$$T_m \frac{d\omega_m(t)}{dt} + \omega_m(t) = K_1 U_a(t) - K_2 M_c(t) \tag{5}$$

In the formula, $T_m = \frac{R_a J_m}{R_a f_m + C_m C_e}$,

$$K_1 = \frac{C_m}{R_a f_m + C_m C_e}, K_2 = \frac{R_a}{R_a f_m + C_m C_e}$$

The motor's transfer function from the control voltage to the angular displacement can be obtained after the LAPLACE conversion:

$$G(s) = \frac{\theta(s)}{U(s)} = \frac{1/C_e}{s(T_m s + 1)} \tag{6}$$

In the formula, $1/C_e$ is the velocity constant, and T_m is the mechanical time constant.

3 Foundation for grey prediction theory

Grey theory was firstly proposed by Professor Deng Julong from Huazhong University of Science and Technology [4]. Through the small amount of incomplete information, the grey differential prediction model was established, thus whitening the grey quantity obtained in the system to improve the control effect and robustness of the controller.

3.1 ESTABLISHMENG OF GM (0, N) MODEL [6]

Supposing $x_1 = \{x_1(1), x_1(2), x_1(3), \dots, x_1(n)\}$ is the dependent variable, $x_i = \{x_i(1), x_i(2), x_i(3), \dots, x_i(n)\}$ is the independent variable, $i = \{1, 2, 3, \dots, N\}$ is the time series data, the step to establish the time series data grey prediction model GM (0, N) with the above small amount of information is:

(1) Initialization time series data

The time series for the dependent variable and independent variable after the initialization is [7]: $x_i^{(0)} = x_i(k) / x_i(1)$

In the formula, $x_i(1)$ is the first value of the variable sequence.

(2) Establish the GM(0, N) model

Supposing $x_i^{(1)}$ is the sequence obtained after one AGO of $x_i^{(0)}$; The prediction model of GM(0,N) can be

obtained as [8]: $x_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(i)}(k) + a$

Parameter $\hat{b} = [b_2, b_3, \dots, b_N, a]$ can obtain the parameter \hat{b} through the least square method:

$$\hat{b} = (B^T B)^{-1} B^T Y \tag{7}$$

In the formula, $B = \begin{bmatrix} x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) & 1 \\ \vdots & & \vdots & \\ x_2^{(1)}(n) & \cdots & x_N^{(1)}(n) & 1 \end{bmatrix}$,

$$Y = \begin{bmatrix} x_1^{(1)}(2) \\ x_1^{(1)}(3) \\ \vdots \\ x_1^{(1)}(n) \end{bmatrix}.$$

The solved approximate time response is:

$$x_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k) + a. \tag{8}$$

Reducing (7) can obtain:

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k), \quad k = 1, 2, 3, \dots, n-1. \tag{9}$$

In the formula $\hat{x}_1^{(0)}(1) = \hat{x}_1^{(1)}(1)$.

3.2 GM(0, N) MODEL PREDICTION PROCESS

GM(0, N) model prediction process is as shown in the following FIGURE 2 [9],

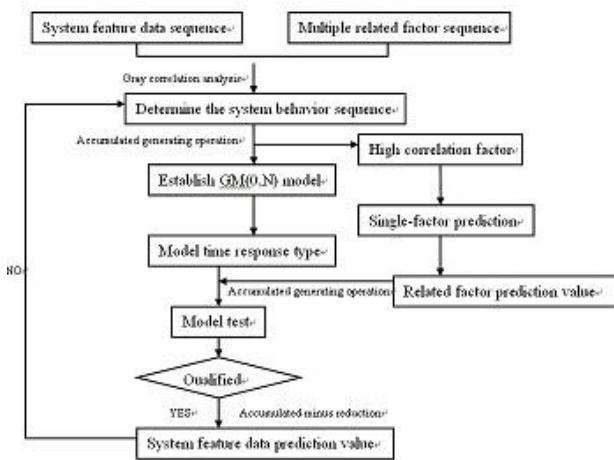


FIGURE 2 Flowchart

4 Design of grey PID controller

4.1 PARAMETER ESTIMATION

If the following nonlinear uncertain system is the research object,

$$\dot{x} = Ax(t) + bu(t) + bD(x, t). \tag{10}$$

$bD(x, t)$ is the uncertain part to meet the system matching condition, and it contains the parameter uncertainty and external interference mentioned previously. Supposing

$$D(x, t) = V_1 x_1 + V_2 x_2 + \cdots + V_n x_n + f(t). \tag{11}$$

When we use the traditional PID control algorithm, the uncertainty existed in the system will influence the system's control performance. In order to reduce the influence brought by the uncertainty, improve the control effect and system robustness, the corresponding GM(0, N) model can be established with previous grey prediction theory, the coarse value of the parameter V in the uncertain part model can be estimated, and the uncertain part $D(x, t)$ can be correspondingly compensated, thus improving the control effectiveness and system robustness. $D(x)$ in $D(x, t)$ cannot be directly measured through our measurement means, therefore, the data obtained through our measurement can be obtained after the grey prediction, and the discrete value is:

$$D(x, k) = \frac{1}{b} (\dot{x}(t) - Ax(t) - bu_p(t)). \tag{12}$$

In the formula, $u_p(t)$ is the control volume of traditional PID.

The specific algorithm for the estimation of coarse value of model parameter V is as follows [9]:

Step 1, Establish the original discrete sequence

$$\begin{aligned} D^{(0)} &= (D(1) D(2) \dots D(N)) \\ f^{(0)} &= (f(1) f(2) \dots f(N)) \\ \left. \begin{aligned} x_1^{(0)} &= (x_1(1) x_1(2) \dots x_1(N)) \\ x_2^{(0)} &= (x_2(1) x_2(2) \dots x_2(N)) \\ \dots \\ x_n^{(0)} &= (x_n(1) x_n(2) \dots x_n(N)) \end{aligned} \right\} \end{aligned} \tag{13}$$

Step 2, Sequence accumulation establishing model

$$x^{(1)} \triangleq \sum_{m=1}^k x^{(0)}(m). \tag{14}$$

Supposing $D^{(1)}$, $f^{(1)}$, $x_i^{(1)}$ ($i=1, 2, \dots, n$) is the accumulating generation sequence of $D^{(0)}$, $f^{(0)}$, x_i ($i=1, 2, \dots, n$). The grey model in the uncertain part $D(x, t)$ is the accumulating generation sequence.

$$D^{(1)}(x, t) = V_1 x_1^{(1)} + V_2 x_2^{(1)} + \dots + V_n x_n^{(1)} + f^{(1)}. \tag{15}$$

Concerning the slowly time-varying interference part, it can be regarded as

$$\begin{aligned}
 f^{(1)}(1) &= f(1) \triangleq f \\
 f^{(1)}(2) &= 2f(1) \triangleq 2f \\
 &\dots \\
 f^{(1)}(N) &= Nf
 \end{aligned}
 \tag{16}$$

Remember the parameter list as

$$\hat{V}^T = (\hat{V}_1 \hat{V}_2 \dots \hat{V}_n \hat{f})^T
 \tag{17}$$

Step 3, Calculation

$$B = \begin{bmatrix} x_1^{(1)}(2) \dots x_n^{(1)}(2) \dots 1 \\ x_1^{(1)}(3) \dots x_n^{(1)}(3) \dots 2 \\ \dots \\ x_1^{(1)}(N) \dots x_n^{(1)}(N) \dots N-1 \end{bmatrix}
 \tag{18}$$

In the formula, $B^T B$ must be reversible. If it is irreversible, N shall be increased or decreased appropriately until $B^T B$ becoming reversible.

Step 4, According to the state $B^T B$ and Formula (15), calculate $D^{(0)}(k)$ discrete sequence,

Step 5, Calculate $D^{(0)}(k)$ accumulating discrete sequence

$$D_N^{(1)} = (D^{(1)}(1) D^{(1)}(2) \dots D^{(1)}(N))
 \tag{19}$$

Step 6, Calculate the uncertain parameter estimating value

$$\begin{aligned}
 \hat{V} &= (B_b^T B_b)^{-1} B_b^T D_N^{(1)} \\
 \hat{V} &= (\hat{V}_1 \hat{V}_2 \dots \hat{V}_n \hat{f})^T
 \end{aligned}
 \tag{20}$$

4.2 GREY PID CONTROLLER

Step 1, Firstly adopt PID control, and the control formula is

$$\begin{aligned}
 u_p &= u(kT) = K_p e(kT) \\
 &+ K_i \sum_{j=0}^k e(jT)T + K_d \frac{e(kT) - e((k-1)T)}{T}
 \end{aligned}
 \tag{21}$$

Step 2, Estimate the parameter V with grey theory. Based on the above control law, estimate the parameter V with compensation control u , the estimator stops working, the grey PID control algorithm is $u+u$, thus inputting the clutch control after the compensation. Wherein,

$$\begin{aligned}
 u_e &= - \left(\sum_{i=1}^n \hat{V}_i x_i + \hat{f} \right) \hat{D}(x, t) \\
 &= \sum_{i=1}^n (V_i - \hat{V}_i) x_i (f(t) - \hat{f})
 \end{aligned}
 \tag{22}$$

5 Numerical simulation

In order to initially verify the PID control algorithm based on grey prediction theory's location tracking efficiency in the DC motor, the system is simulated with MATLAB simulating software. According to the selected related parameter of DC motor in Formula (6), the motor's transfer function is:

$$G(s) = \frac{\theta(s)}{U(s)} = \frac{1/C_e}{s(T_m s + 1)} = \frac{86}{s(0.01s + 1)}
 \tag{23}$$

In the simulation process, the specific simulation parameter is as follows: $V1=[5 \ 5]$, $f=5$, $k_p=30$, $k_d=5.0$. The simulation result is as shown in FIGURE 3-8, Figure 3-5 are the location tracking under the traditional PID control, error and controller output situation respectively, Figure 6-8 are location tracking under the grey prediction PID control, error and controller output situation respectively.

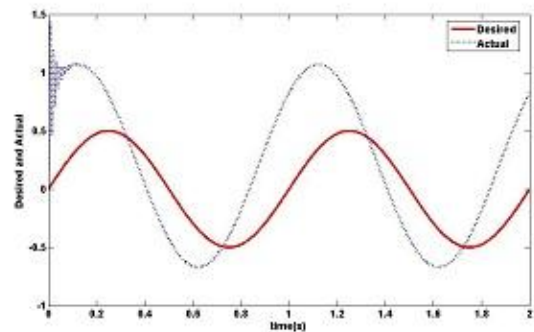


FIGURE 3 Traditional PID control location tracking

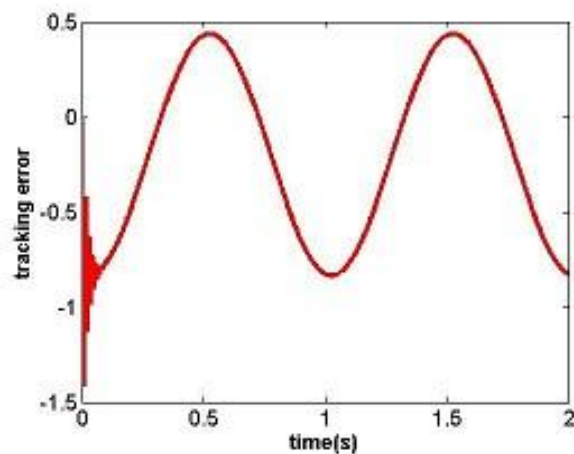


FIGURE 4 Traditional PID control location tracking error

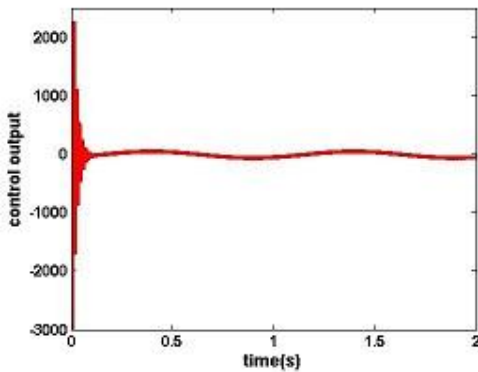


FIGURE 5 Traditional PID control output

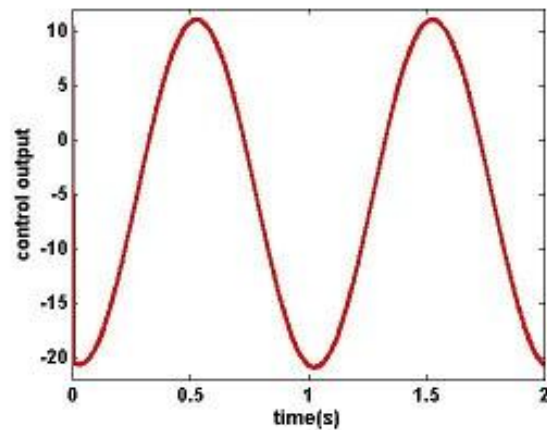


FIGURE 8 Grey PID control output

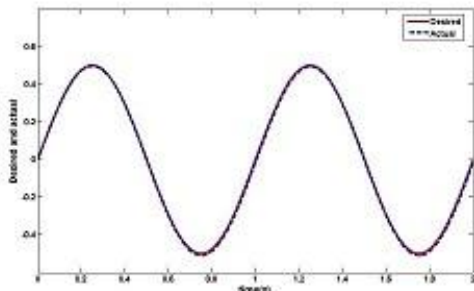


FIGURE 6 Grey PID control location tracking

According to the above simulation result, it can be seen that under the same simulation environment and parameter condition, the PID control algorithm based on grey prediction theory showed a good tracking performance in the location tracking process in DC motor. Compared with the traditional PID control algorithm, PID control algorithm based on grey prediction theory can effectively use the grey theory's processing capability on the unknown information data. The uncertain amount (parameter uncertainty and disturbance) in the system is established the grey model and effectively predict and compensate the unmodelled feature and disturbance signal in DC motor servo system, thus improving the control accuracy of the controller.

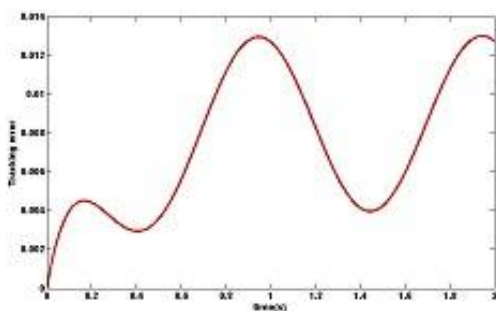


FIGURE 7 Grey PID control location tracking error

6 Conclusions

Concerning the uncertainties may be existed in DC motor servo system and the control quality that the external disturbance may influence the control algorithm of traditional PID, a kind of PID control algorithm combined the grey prediction theory and traditional PID control algorithm was proposed and applied in DC motor servo system. With the grey theory's processing ability on the unknown information data, the proposed control algorithm established the grey model for the uncertainties, and real-time compensated the system's unmodelled feature and disturbance signal. The simulation result can show that under the same simulation environment and parameter condition, compared with the traditional PID control algorithm, the proposed PID control algorithm based on grey prediction theory showed a good tracking performance in the location tracking process of DC motor. It can effectively use grey theory's processing capacity on the unknown information data. The uncertain amount (parameter uncertainty and disturbance) in the system is established the grey model and effectively predict and compensate the unmodelled feature and disturbance signal in DC motor servo system, thus improving the control accuracy of the controller.

Acknowledgments



This work is funded by the Scientific Research Projects of Xianyang Normal University (No.11XSYK 329, No.13XSYK058).

References

[1] Hou Li, Fan Qingwen, Huang Chengxiang 2003 *Design of Mechanical & Mechanical-Electrical Integration System* Press: Beijing: Higher Education Press (In Chinese)

[2] Zhang Zhihui, Li Yuren 2007 Application of Kalman Filtering Technique in Servo System *Small & Special Electrical Machines* 35(12) 16-8 (In Chinese)
 [3] Jingdong Zhang, Rui Tang, Yongqiao Wei 2013 Fuzzy Adaptive PID Control for Parallel Machine Tool *Applied Mechanics and Materials* 273 660-4 (In Chinese)
 [4] Li Feng, Deng Julong 2000 Development Overview of Grey System Theory *Information and Development* 3 6-9 (In Chinese)

- [5] Zhang Jihe 2008 *Motor and Control* Press: Southwest Jiaotong University (*In Chinese*)
- [6] Deng Julong 1990 *Grey System Theory Tutorial* Press: Huazhong University of Technology (*In Chinese*)
- [7] Ji Wenjuan, Gu Wenben 2006 Application of GM(0,N) Grey Prediction Model in the Prediction of Yunnan Late Autumn Crop Yield *Chinese Agricultural Meteorology* 27(3) 229-32
- [8] Xu Rufu, Wang Wenchang, Yi Dong, Zhang Wei, Yin Quanhuan 1999 Time Series Data GM(0,N) Prediction Model and Application *Chinese Health Statistics* 16(3) 162-3 (*In Chinese*)
- [9] Yan Yongfei, Zhu Shunying, Wang Hong, Liu Jianting 2010 Highway Transport Volume Prediction Method Based on GM(0,N) Gray System *Wuhan University of Technology Journal* (Transportation Science & Engineering Edition) 34(1) 93-6 (*In Chinese*)

Authors	
	<p>Hong Li</p> <p>Current positions, grade: working in the School of Information Engineering, Xianyang Normal University</p> <p>University studies: Master degree in computer applications technology from Northwest University for Nationalities in 2006.</p> <p>Publications: authored or co-authored more than 20 papers.</p> <p>Scientific interests: different aspects of control engineering, image processing and artificial intelligence.</p>
	<p>ZongZhe Wang</p> <p>Current positions, grade: an engineer</p> <p>University studies: received the Bachelor degree in electrical automation from Xi'an University of Technology.</p> <p>Scientific interests: control engineering and electrical automation.</p>

A new automatic registration method for InSAR image based on multi-step strategy

Bingqian Chen*

School of Environment Science and Spatial Informatics, China University of Mining and Technology, 1 Daxue Road, Xuzhou, 221116, China

Received 12 May 2014, www.tsi.lv

Abstract

Interferometric Synthetic Aperture Radar (InSAR) technology has been widely used in various applications. The registration of SAR (Synthetic Aperture Radar) images is the first step in interferometric processing therefore accurate registration is essential for the successful creation and interpretation of interferometric products. However, with the growing number of SAR satellite launch and the amount of data acquisition, the degree of automation of image registration have become increasingly demanding. In this paper, we propose an automatic registration approach based on multi-step matching strategy. In the first step, key points are detected and matched using modified scale invariant feature transform (SIFT) operator which modified by us reducing the influence of speckles. In this step, owing to the existing of speckle and the defect of matching strategy of SIFT operator, the expected level of matching accuracy is about 2 to 3 pixels. In the second step, correlation matching (CM) is used to exclude the matched points with low correlation. In the third step, the probability relaxation (PR) algorithm based on global matching is used to induce consistency constraint and ensure reliability of the matching result. Finally, corresponding transformation function is determined through the relationship established by matched point pairs. In order to verify the applicability of proposed methodology, two SAR images acquired over mountainous regions are used in our experiment. The experiment results show that subpixel registration accuracy and good efficiency have been achieved, which demonstrates the correctness and feasibility of proposed method.

Keywords: InSAR, Image registration, Feature detection, SIFT, Correlation matching

1 Introduction

InSAR is a technology developed in recent decades, which utilizes the relationship between the phase difference and spatial distance difference of complex data obtained in two repeat observations to extract three-dimensional information or elevation change information of ground surface [1, 2]. Due to its outstanding advantages of fast, high-precision, all-time, all-weather and large areas, it has been used in many domains, like Digital Elevation Model (DEM) generation[3], target recognition [4], volcano activity detection [5], landslide movements studies [6], and many more [7-9]. Prior to most of mentioned applications, SAR images are required to be aligned accurately (usually subpixel level).

The definition of the image registration can be explained as two images obtained from different time, sensors, or perspectives are unified to the same coordinate system according to certain criterion. In general, most image registration methods consist of three main steps: locating of GCPs (Ground Control Points), matching of GCPs, determining of geometrical transformation function. Some obvious ground objects can be used as GCPs, like buildings, bridges, crossroads, etc. The identification of sufficient GCPs pairs allows the transformation function to be determined. Finding suitable GCPs pairs is the most difficult step.

Traditional registration method makes use of GCPs determined by operating staff, which is time-consuming and imprecision and is practically impossible for single look complex (SLC) SAR data. Therefore, many researchers have focused on the study of automatic registration method [10-11]. However, the complexity of registration of SAR image is much higher than optical image due to the existing of speckle noise and blur-texture feature.

In our study, we introduce feature-based scale invariant feature transform (SIFT) operator to extract prominent features points and complete initial registration, then area-based correlation matching (CM) is used to exclude the matched points with low correlation, to achieve the effect of removing of mismatch points. Furthermore, the probability relaxation (PR) technique based on global matching algorithm is used to induce consistency constraint and ensure reliability of the matching result. Finally, the parameters of transformation function are computed through the established key point correspondence, and slave image is transformed through the transformation function. The proposed method is an automatic approach using a multistep matching strategy, which completes the registration procedure from coarse to fine precision. The results of the experiment are promising in their reliability, accuracy, and computing efficiency, which prove the feasibility of proposed method.

* *Corresponding author* e-mail: bqccumt@gmail.com

2 Methodology

Our method mainly contains four steps. The matching accuracy of key points is improved through the first three steps, moving one-step at a time, coarse to fine, thus an accurate mapping function model is established through the fourth step. The flowchart of the main steps of proposed method is illustrated in Figure 1. In the first step, key points are detected and matched using modified SIFT operator. In this step, due to the existing of speckle and the defect of matching strategy of SIFT operator, the expected level of matching accuracy is about 2 to 3 pixels. In the second step, CM is used to exclude the matched points with low correlation, to achieve the effect of removing of mismatch points. In the third step, the PR technique based on global matching algorithm is used to induce consistency constraint and ensure reliability of the matching result. Finally, the parameters of transformation function are computed through the established key point correspondence, and slave image is transformed by means of the transformation function. Although the slave images are SLC data, the whole procedure processed intensity images or power-detected images, i.e., it avoided computing the complex data.

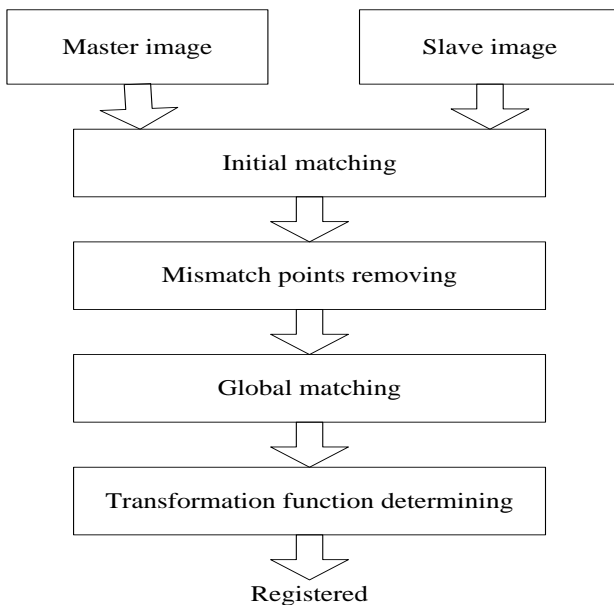


FIGURE 1 Flowchart of the main steps of proposed model

2.1. INITIAL MATCHING

2.2.1 Key points detection

The entire process flow of proposed method starts from key point detection. For this purpose, the first step is to establish the image scale space. For a two-dimensional

image, its corresponding scale space $L(x, y, \sigma)$ can be obtained by convolving the image with Gaussian kernel:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (1)$$

where $I(x, y)$ represents the two-dimensional image coordinate, $G(x, y, \sigma)$ represents the Gaussian kernel, σ represents the variance of the Gaussian distribution. In order to detect key points efficiently in the scale space, difference of Gaussians (DoG) scale space $D(x, y, \sigma)$ is created by means of convolving the image with different scale space Gaussian kernels in the second step, and DoG scale space is defined as:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ = L(x, y, k\sigma) - L(x, y, \sigma), \quad (2)$$

where k is the scale difference factor. Finally, the extrema are obtained through comparing the value of DOG of sample pixel with its surrounding eight pixels in the same scale space and the eighteen pixels in adjacent upper and lower scale space, respectively. If the value of sample pixel is larger or smaller than other neighbours, it will be accepted as an initial key point.

Here considering the requirement of computational efficiency and the influence of speckle noise, the original algorithm of key point detection is adapted which skip the searching for key points in first octave of the scale-space pyramid. This modification is called SIFT-OCT, more details about SIFT-OCT can be found in literature [12].

2.1.2 Key Points Matching

According to the original matching strategy of SIFT operator proposed by Lowe [13], the smallest Euclidean distance criterion is adopted as similarity index, in other words, if the distance between two descriptor vectors defined by SIFT operator are smaller than others, corresponding points are taken as a matching pair. In our initial matching process, the strategy of smallest Euclidean distance is adopted. However, the quality of matching result using simple distance matching strategy is poor and cannot satisfy our requirement of accurate registration. An example of original SIFT operator matching result is illustrated in Figure 2. As shown in Figure 2, the positions of dots in the left image correspond to the position of dots in the right image and mismatch situations happening in the same colour rectangle frame, which will affect the final accuracy of image registration, therefore, in the following steps, the matched pairs will be refined by CM and Global image matching.

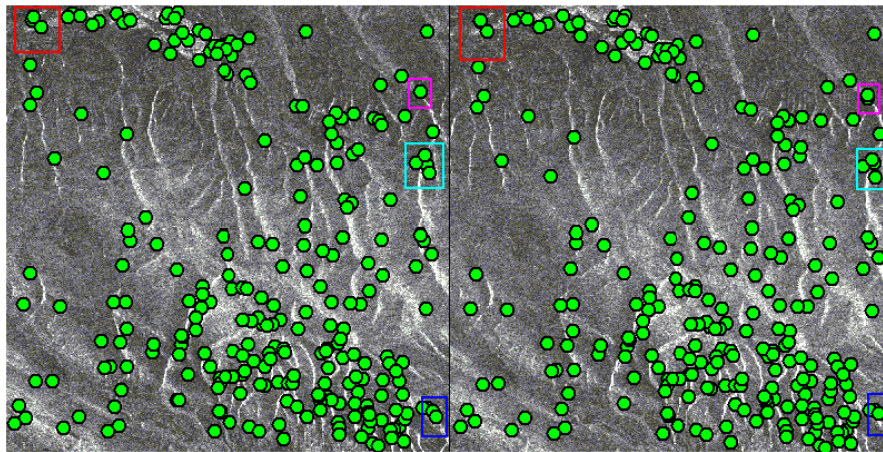


FIGURE 2 The matching result by original SIFT operator

2.2 MISMATCH POINTS REMOVING

Area-based CM methods also called template matching [14]. A certain number of feature points are firstly selected based on some kind algorithm (e.g. SIFT operator) in the master image, and then a certain size window with the feature point centred is determined, and the window moves on the slave image pixel by pixel to search the window with high set correlation. Finally, the central pixels of the two windows with highest correlation are considered as the matching points. Generally speaking, the conventional CM methods include correlation coefficient method, Fourier method and mutual information method. Nevertheless, the calculation of the mentioned methods is complex especially when deal with large images. Because we have already obtained initial matching point pairs by SIFT operator in section 2.1 and our ultimate purpose of the study is the removing of mismatch points ,therefore, a search in all over the image is unnecessary we can calculate the correlation value of the matched points directly and judge whether it meet the requirement of threshold value. If not, these points will be removed.

Correlation coefficient, is originally defined by complex calculation and now is defined by us just calculating intensity value, which avoids computing the complex data thus computational efficiency is improved. Here, correlation coefficient is defined as follows:

$$\rho = \frac{\sum_{m=1}^k \sum_{n=1}^g R(m, n)S(m, n)}{\sqrt{\sum_{m=1}^k \sum_{n=1}^g R(m, n)^2} \sqrt{\sum_{m=1}^k \sum_{n=1}^g S(m, n)^2}}, \quad (3)$$

where ρ is the correlation coefficient (between 0 and 1), $R(m, n)$ and $S(m, n)$ represent the coordinate of certain matched point in master image and the corresponding coordinate of point in slave image, respectively. The matching window size is $k \times g$. The detailed implementation steps of CM are described as follows:

(i) The acquirement of initial matching points: The initial matching points are obtained by modified SIFT operator, as mentioned in section 2.1. We can assume set $R' = \{R_1', R_2', \dots, R_i'\}$ is initial matching points in the master image and $S' = \{S_1', S_2', \dots, S_i'\}$ is the corresponding matching points in the slave image, respectively.

(ii) The calculation of correlation coefficient: The value of correlation coefficient of the matched points is calculated according to equation (3), and a certain size search window is needed in the calculation processing of correlation coefficient. From section 2.1 we can know that the extrema are detected in the DoG scale space by comparing every pixel to the eight neighbouring pixels in the same scale space, therefore, we can choose the range of 3×3 as our size of search window.

(iii) The judgement of correlation coefficient: If the correlation coefficient ρ is greater than threshold Y , it is accepted as right matched points and retained, otherwise excluded. Here 0.3 is adopted as the threshold Y in our experiments.

Some obvious mismatch points can be removed by CM, and the number of initial matching points also can be lessened to a certain extent by adjusting the threshold value of correlation coefficient, which will reduce the amount of calculation of next steps. However, CM belongs to isolated single point matching and does not take account of the compatibility of the matching result of one point pair with its neighbours, which leads to the uncertainty existing of matching result and false matches may still occur. Therefore, PR algorithm based on global matching is introduced in the next step.

2.3 GLOBAL MATCHING

In this section, the PR algorithm is adopted to further refine and ensure reliability of the matching result obtained in section 2.2. PR algorithm was firstly used for image registration by Rosenfeld et al. in 1976 [15]. The PR algorithm is a parallel and iterative computation process. Every point on the processing is in parallel in every iteration process, and the iteration results of the

point are adjusted by means of the result of surrounding points in next iteration. This parallel iterative process not only taken into account the impact on the surrounding results, but also has the characteristics of parallel processing, therefore, it has been widely used in various fields. In our study, the PR algorithm is used and two matched point sets obtained in section 2.2 are regarded "object" and "class", respectively. Therefore, the problem from registration converts into classification. The detailed process is described as follows.

Since we have acquired preliminary matching point pairs in section 2.2, here we assume matched points set $T = \{T_1, T_2, \dots, T_j\}$ as "object" set and correspondign set $S = \{S_1, S_2, \dots, S_i\}$ as "class" set. As shown in Figure 3, pixel m is the "object" in the master image, pixel n is the corresponding "class" in the slave image. And now the image matching become solving the problem $T_m \in S_n$. To improve the reliability of result of classification, the global consistency must be considered, i.e. classification results are in harmony with each other. Therefore, on the basis of the requirement of PR algorithm, the probability P_{mn} of $T_m \in S_n$ and the compatible coefficient $C(m, n; t, g)$ of $T_m \in S_n \cap T_t \in S_g$ should be determined firstly.

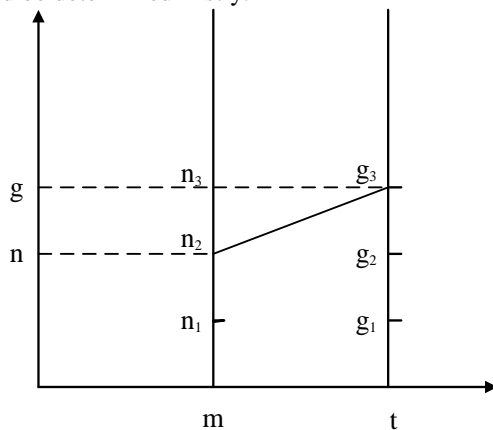


FIGURE 3 PR algorithm with bridge model

Because the coherence of matched point pairs are calculated by CM algorithm in section 2.2, here we take the correlation coefficient ρ_{mn} as the initial value P_{mn} of $T_m \in S_n$, and also we define the correlation coefficient $\rho(\bar{m}, \bar{t}; \bar{n}, \bar{g})$ between interval $[m, t]$ and $[n, g]$ (Figure 3) as the initial compatible coefficient $C(m, n; t, g)$ of $T_m \in S_n \cap T_t \in S_g$, i.e.,

$$C(m, n; t, g) \propto \rho(\bar{m}, \bar{t}; \bar{n}, \bar{g}). \tag{4}$$

While the initial measurement of P_{mn} and $C(m, n; t, g)$ are determined, the relaxation iterative procedure can be conducted according to the following equations:

$$Q(m, n) = \sum_{t=1}^{j(M)} \left(\sum_{g=1}^{i(N)} C(m, n; t, g) P(m, n) \right) \tag{5}$$

$$P^{(u)}(m, n) = \frac{P^{(u-1)}(m, n)(1 + IQ(m, n))}{\sum_{n=0}^{i(F)} P^{(u)}(m, n)}$$

where $j(M)$ represents the number of neighbor points of point m , $i(N)$ and $i(F)$ represent the number of candidate matching points of slave image, u is the ordinal number of iteration, I represents the incremental coefficient. Iteration termination condition of equation 5 is defined as:

$$|P^{(u)}(m, n) - P^{(u-1)}(m, n)| \leq \varepsilon. \tag{6}$$

3 Evaluation of Proposed Model Performance

To evaluate the performance of proposed model, several criteria are introduced and described as follows:

(i) Running time: For testing the operating efficiency of proposed method, we recorded the turnaround time, which includes the time for the detection and execution of the matching process.

(ii) Standard deviations (STDs): To assess the performance and verify the feasibility of the proposed method, the STDs are introduced as an indicator of measuring the accuracy of the registration. The STDs computing formula is defined as follows:

$$\left\{ \begin{aligned} STD_{range} &= \sqrt{\frac{\sum_{i=1}^n (x_i - x'_i)^2}{n-1}}, \\ STD_{azimuth} &= \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n-1}} \end{aligned} \right. \tag{7}$$

where the (x_i, y_i) represents the coordinate of pixel i in master image, (x'_i, y'_i) represents the coordinate of the same object in the transformed slave image, n is the number of matched point pairs.

(iii) Interferogram quality evaluation: Interferogram is the product obtained by complex conjugate multiplying of two registered images, and its fringes definition is taken as the criterion of registration quality by many researchers [16]. That is, if the interferometric fringes are clear, it indicates a good registration result is achieved. However, it is a sufficient condition, not a necessary condition, i.e. although clear interferometric fringes indicate a good quality of registration, the blurred fringes in interferogram not always mean poor registration accuracy because the quality of interferogram depends on not only registration but also temporal baseline, spatial baseline and noise. Here the quality of an interferogram is only used as qualitative description of registration results.

(iv) Coherence estimation: Coherence is an important concept in the InSAR technology, which indicates the degree of similarity between the two SAR images. The coherence value is not only an important evaluation criterion for the quality of local interferometric fringes, but also a necessary condition to guarantee the reliability of the results [17]. In our study, the mean value and corresponding STDs of coherence are used to verify the reliability of registration result. Here the coherence value $\gamma(0 < \gamma < 1)$ is defined as follows:

$$\gamma = \frac{|E[f_1 \cdot f_2^*]|}{\sqrt{E[|f_1|^2]E[|f_2|^2]}}$$
(8)

where f_1 and f_2 represent the two registered SAR images.

4 Experiments and Results

4.1 DATASET

To verify the feasibility of the proposed method, two images (size 600×600 pixel) acquired by X-band TerraSAR-X satellite (launched by Deutsches Zentrum für Luft- und Raumfahrt in 2007) are used in our first experiment. The images were acquired over Van, Turkey, featuring a mountain land cover class. From Table 1 we can know that the time difference between two images is eleven days.

TABLE 1 Details of TerraSAR-X images

Image	TerraSAR-X (Master image)	TerraSAR-X (Slave image)
Mode	Stripmap Mode	Stripmap Mode
Pixel Spacing (m)	3	3
Incidence Angle	33.18°	33.18°
Date of Acquisition	31/10/2011	11/11/2011
Wavelength (cm)	3.1	

4.2 EXPERIMENT RESULTS

The matching results using proposed method in different steps are seen in Figure 5(a), (b), and (c). The number of initial key points detected by modified SIFT operator is 2266 in the master image and 1951 in the corresponding slave image (Table 2). After applying of SIFT matching strategy 264 matching point pairs are obtained. In this step, owing to the speckle existing and the defect of matching strategy of SIFT operator, mismatch occasions may occur (Figure 4(a)), which lead to the high STDs in range and azimuth directions (Table 2). So in the next steps, CM and PR matching are adopted to refine the matched pairs, the matching results are seen in Figure 4(b) and (c), respectively. From Figure 4(b) we can know that the quality of matched points is improved but still mismatching cases happening. Figure 4(c) shows that the occasions of mismatching are disappeared and the final

number of matching point is reduced to 75 and the corresponding STDs are subpixel in both range and azimuth directions (Table 2). The improvement in the STDs show that initial matching by SIFT operator and CM provides reliable initial locations.

To further examine the effect of image registration, an interferogram (Figure 5) is produced by a point-wise complex conjugate multiplication of corresponding pixels in master image and registered image. And also, corresponding coherence map (Figure 6) is obtained by calculating equation 8.

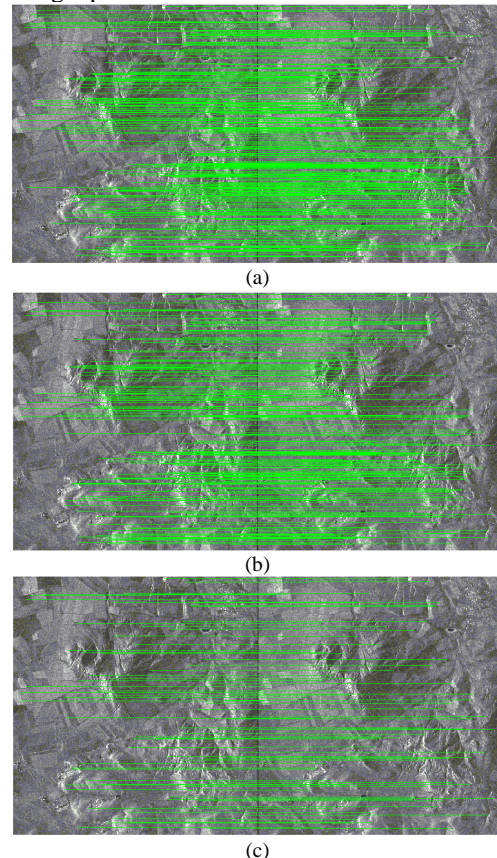


FIGURE 4 Matching results in different steps. (a) Matching result using the SIFT operator. (b) Matching result after using CM to (a). (c) Matching result after using PR matching to (b)

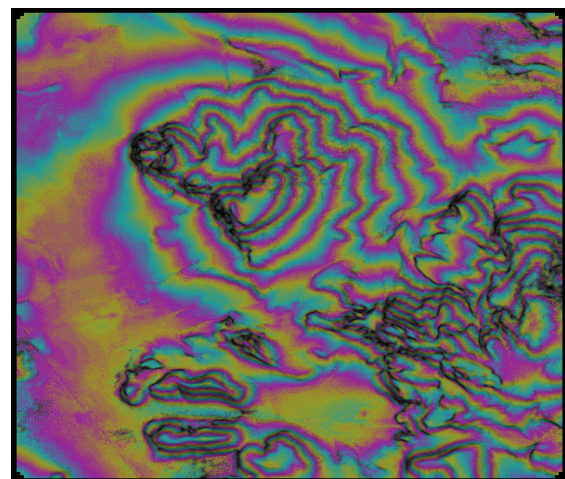


FIGURE 5 Interferometric fringes (Van, Turkey) of X-band images

By visual inspection, we can find that most of the interferometric fringes are clear and continuous, which indicates the good quality of interferogram (Figure 5), and corresponding a good coherence map (Figure 6(a), high coherence is found in bright areas and low coherence in dark areas.). In order to describe the value of coherence quantitatively, the histogram of coherence image is introduced (Figure 6(b)) and the mean of the

distribution and corresponding STDs are used to coherence estimation. In Figure 6(b), the mean is 0.8395, while its STDs is 0.0454. Both the high quality of interferogram and coherence map also proves the good result of registration. And also from Table 2 it can be observed that the process speed of the test data is near real time and sufficient for our work requirement.

TABLE 2 Experimental results for TerraSAR-X images

Matching Algorithm	SIFT	SIFT + CM	SIFT+CM+PR
Number of key points	Ref:2266 Inp:1951	Ref:264 Inp:264	Ref:191 Inp:191
Number of matches	264	191	75
STDs (pixel)	Range:2.605 Azimuth:2.413	Range:1.217 Azimuth:1.130	Range:0.0504 Azimuth:0.0717
Turnaround time (s)	13		

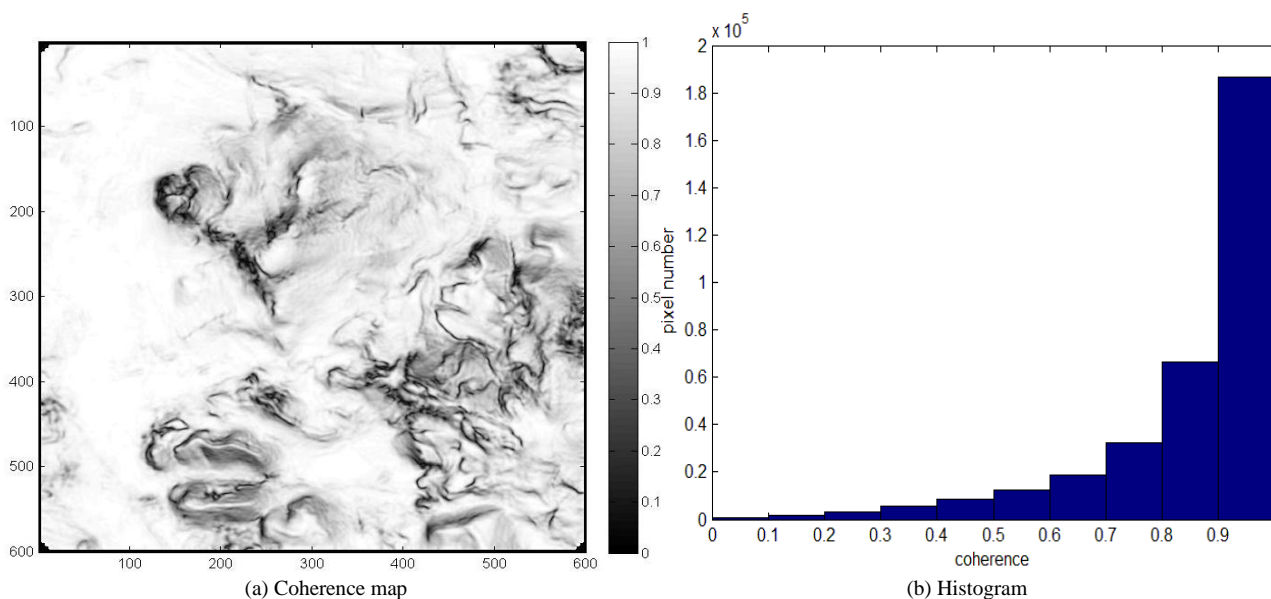


FIGURE 6 Coherence map and the corresponding histogram of coherence

5 Conclusions

In this paper, we present a novel SAR image registration method based on multi-step, which completes the registration procedure from coarse to fine precision. The feature-based SIFT operator is firstly introduced to extract prominent features points and complete initial matching, then area-based CM is used to exclude the matched points with low correlation, to achieve the effect of removing of mismatch points. Finally, the PR algorithm based on global matching is used to induce consistency constraint and ensure reliability of the matching result. The proposed method is an automatic approach which completes the entire registration procedure without manual intervention.

We conduct one experiment using the proposed method, and make a detailed analysis to the experiment results in many aspects, i.e. key point detection, quality of interferogram and coherence map, matching accuracy and process turn around time. The experiment results

show that the proposed methodology shows robust performance and subpixel registration accuracy can be achieved. The experiments results have also shown the process speed of the test data is near real time, which indicate the process efficiency can be guaranteed for larger images. Moreover, another advantage of our method is that it does not require special filtering process before the SAR image registration.

Acknowledgments

This work is funded by the natural science foundation of china (No.41071273 and No. 41272389), a project funded by the priority academic program development of Jiangsu Higher Education Institutions (No.SZBF2011-6-B35), A project of graduate research and innovation of ordinary university in Jiangsu Province (No.CXZZ13_0936). Relevant radar data were provided by the German Aerospace Centre TerraSAR-X Science Plan (LAN1425).

References

- [1] Ouchi K 2013 *Remote Sens.* **5** 716-807
- [2] Burgmann R, Rosen P A, Fielding E J 2000 *Annu. Rev. Earth Planet. Sci.* **28**(1) 169-209
- [3] Ortiz S M, Breidenbach J, Knuth R, Kändler G 2012 *Remote Sens.* **4**(3) 661-81
- [4] Huan R H, Pan Y 2013 *Prog. Electromagn. Res.* **134** 267-88
- [5] Papageorgiou E, Fomelis M, Parcharidis I 2012 *IEEE J-STARS.* **5**(5) 1531-7
- [6] Liu P, Li Z H, Hoey T, Kincal C, Zhang J F, Zeng Q M, Muller J P 2013 *Int. J. Appl. Earth Obs. Geoinf.* **21** 253-64
- [7] Kobayashi S, Omura Y, Sanga-Ngoie K, Widyorini R, Kawai S, Supriadi B, Yamaguchi Y 2012 *Remote Sens.* **4**(10) 3058-77
- [8] Pivot C F 2012 *Remote Sens.* **4**(7) 2133-55
- [9] Gama F F, Santos J R D, Mura J C 2010 *Remote Sens.* **2**(4) 939-56
- [10] Bradley P E, Jutzi B 2011 *Remote Sens.* **3**(9) 2076-88
- [11] Ben-Dor E, Brook A 2011 *Remote Sens.* **3**(1) 65-82
- [12] Schwind P, Suri S, Reinartz P, Siebert A 2010 *Int. J. Remote Sens.* **31**(8) 1959-80
- [13] Lowe D G 1999 Object recognition from local scale-invariant features *In Processing of International Conference on Computer Vision (Corfu, Greece)* 1150-57
- [14] Chen J H, Chen C S, Chen Y S 2003 *IEEE Trans. Signal Process.* **51**(1) 230-43
- [15] Rosenfeld A, Hummel R A, Zueker S W 1976 *IEEE Trans. Syst. Man Cybern.* **SMC-6** 420-33
- [16] Eldhuset K, Aanvik F, Aksnes K 1996 *First results from ERS tandem InSAR processing on Svalbard* <http://www.geo.unizh.ch/rsl/fringe96/papers/eldhuset-et-al/> (accessed 8 January 2008).
- [17] Gens R 1998 *Quality assessment of SAR interferometric data* Hanover University: Hanover, Germany

Authors



Bingqian Chen, born on September 24, 1986, China

University studies: Doctoral candidate

Scientific interest: His research interests include radar data processing, optimization algorithms, data fusion, and use of remote sensing information for Civil Protection applications.

Experience: Bingqian Chen was born in Henan Province, China. He received the B.S. degree in surveying and mapping engineering at the Henan Polytechnic University in June 2010. He has been studying the postgraduate program for Doctor degree on Geodesy and Surveying Engineering in the School of Environment Science & Spatial Informatics at China University of Mining & Technology since September, 2012.

An approach for reference model implementation by predicting all possible output of design

Lirong Qiu*

School of Information Engineering, Minzu University of China, Beijing, China

Received 12 May 2014, www.tsi.lv

Abstract

In verification system, it is preferable to build reference model at transaction level which does not produce the output as the same latency as the design itself. But due to the lack of accurately modelling design's behaviour, there are some scenarios that design's output is different with reference model's output due to the different processing delay of stimulus. Scoreboard can get lots of comparison failure when it tries to do comparison between the output of reference model and design under such scenarios. In this case, neither reference nor design is wrong from functionality, but output comparison failure will mix up with the true design issue and bring trouble to the automatic check on design's behaviour. Cycle based reference model does not have such problem. But it usually takes great effort to implement cycle based reference models and maintain them. This paper provides its study on implementation style of reference model. By predicting all possible output of design, this paper presents a method for reference model to handle such stimulus competition scenarios at the transaction level. The paper also discusses the reference model's reaction effect on generator, which helps the test hit design's corner case.

Keywords: System Verilog, reference model, scoreboard, competition stimulus, coverage driven verification

1 Introduction

Verification usually consumes about 70% of the IC design's effort [1]. A lot of verification methodologies are proposed in recent years. Constrained-random stimulus is one the most important principles in IC verification methodologies [2].

Random base stimulus can only be generated automatically. For automatically generated stimulus, reference model or a scoreboard will be used to predict the results of the stimulus and compare these results with output of design in an automated way. Figure 1 shows the infrastructure of verification system with reference model. As illustrated in the Figure 1, the reference model and the design under verification are subjected to the same stimulus and their output is compared for discrepancies.

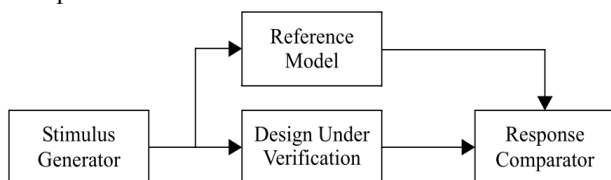


FIGURE 1 Common architecture of verification system

A reference model can be implemented at three levels of precision [3].

Reference models can provide transaction-level functionality.

Reference models can also be cycle accurate.

Reference models can provide rough justification by checking the validity of the DUT behaviour given some input and output. The DUT internals might be used for justification.

1.1 THE ISSUE OF TRANSACTION LEVEL REFERENCE MODEL

The transaction level functionality is the most commonly used reference model, so reference models are usually implemented with C, C++ and System C languages [1, 5]. It is thought that by using a common language the design and verification can proceed smoothly from system-level and architectural-specification down to detailed implementation [7]. However due to the lack of accurately modelling design's behaviour, some scenarios may make reference model and design have different output. Neither reference nor design is wrong from functionality under such circumstance, but output comparison failure will mix up with the true design issue and bring trouble to the automatic check on design's behaviour.

1.2 AN EXAMPLE WHICH HAS DIFFERENT OUTPUT BASED ON THE DELAY OF STIMULUS PROCESSING

Figure 2 shows the infrastructure of a simple design under verification and as followed its normally implemented reference model with C language. In this

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

example the design has 3 interfaces. Interface A is a port for data input. The input is a kind of packet with header, payload and CRC checksum. Interface B is a control interface. Controlling commands for design are injected by interface B. Interface C is for output. Design's behaviour is quite simple: it checks the input packets, discards the packets with CRC error and saves the payload of CRC correct packets into a local buffer. When a flush command is received from interface B, design will output the packets' payload in its local buffer through interface C. If local buffer is full, the input packets will be discarded.

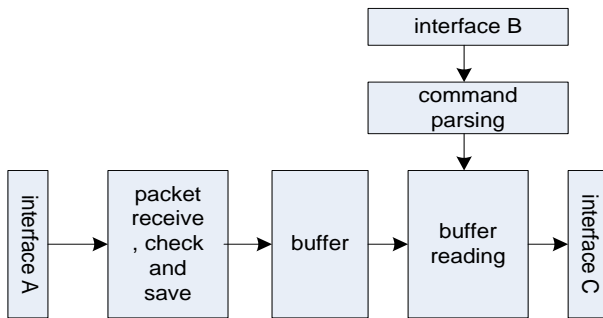


FIGURE 2 An example of design's infrastructure

Reference model for the example design is shown:

```

#include <stdio>
#define BUF_LENGTH xx
typedef struct tagPacket {
    char header[4];
    char payload[1000];
    char crc[4];
    unsigned int length;
} MyPacket;
MyPacket Buffer[BUF_LENGTH];
unsigned int ValidPacketNum;
void flush(MyPacket* output, valid_length) {
    Output = Buffer;
    Valid_length = ValidPacketNum;
    ValidPacketNum = 0;
}
void packet_receive(MyPacket input) {
    Int i = 0;
    If(ValidPacketNum== BUF_LENGTH){
        printf("buffer is full, packet is discarded\n");
        return;
    }
    If(!crc_check(input) {
        Printf(" CRC error is found, packet is discarded\n");
        return;
    }
    for(i=0;i<4;i++)
        Buffer[ValidPacketNum].header[i]= input.header[i];
    for(i=0;i<input.length;i++)
        Buffer[ValidPacketNum].payload[i]= input.payload[i];
    for(i=0;i<4;i++)
        Buffer[ValidPacketNum].crc[i] = input.crc[i];
    Buffer[ValidPacketNum].length = input.length;
    ValidPacketNum++;
}
  
```

Reference model is formed by two functions and the global variables which act as the local buffer and the valid packet number saved in the buffer. One function is to process the packets data and the other function is to process the command. As reference model is implemented at transaction level, it is not necessary to emulate the exact function of design. The data saved in the buffer can be packets type and it is not necessary to get packet header and appended CRC removed. The transaction level reference model is easy to be implemented and integrated from system level down to detailed implementation. Moreover, the simplification makes transaction level model has less bug embedded and can be entitle more confidence as a real golden model.

However, there is no timing delay for reference model to run the function task such as packet receiving, packet checking, command executing and buffer flushing. Due to the lack of timing delay, reference model may predict a different output as design does for the case that buffer is full and a flush command is coming shortly after another packet injection:

For reference model, the packet will be definitely discarded as buffer is full and flush command is not injected yet.

For design, the result depends on the delay of processing the packet and flushing command. For example, if design's behaviour is like this: when flush command is going on and the left byte is available for the new valid incoming packet, the incoming packet will be saved (not be flushed out. Chapter 4.3 will discuss the case of this packet's flushing out). Alternatively, it will be discarded.

Therefore, for real design, we get two different results based on stimuli's different processing time. If result is first one, it will be different with reference model's output. Under such circumstance we cannot say either design or reference model is wrong, because they both behave rightly according to the functionality. The most outstanding character of such scenarios is that the stimuli to be processed have competition. Who is the winner decides the processed result. In this paper, we call such scenarios as stimulus competition scenarios. In addition, the bottom of this issue is that reference model cannot process the packet or command in the same synchronization step as design. It is usually called as asynchronous issue between reference model and RTL. Stimulus competition scenarios in a verification system and asynchronous issue between reference models and designs commonly exist. Moreover, scenario of buffer flushing and packet processing under buffer full condition is also an important corner case should be covered by verification system. Although we do not care if the packet is discarded or saved, we do care if design will not hang up. This asynchronous issue must be worked around for our test target.

In fact, to work around this asynchronous issue between reference model and design, we have several choices to do: Use reference model with accurate timing

delay. Alternatively, make dedicated tests to test such scenarios and handle the output comparison specially. However, of them is perfect and neither both of them need more effort on reference model or test suites. This paper provides its study on implementation style of reference model. By predicting all possible output of design, this paper presents a method, which entitles reference model the ability of handling the timing sensitive scenarios automatically at the transaction level for testbench. The following of this paper includes:

How to implement reference model and handle stimuli competition scenarios at transaction level are discussed and an example is introduced based on the implementation method.

Several factors are introduced to optimize this method.

Moreover good test program need to provide more automation to maximize the functional coverage from each test case and reduce the time needed to create a test case [4 and 6]. So it is better to let verification system can handle such asynchronous issue automatically and hit such corner case easily. Chapter 4.2 also discuss an advantage about the method to implement reference model.

2 The approach to handle stimulus competition scenarios at transaction level

In previous chapter, we have pointed out that due to asynchronous issue, stimulus competition scenarios may make reference model and design have different output. The main reason is transaction level reference usually takes no simulation time to process a stimulus while RTL design cannot. Cycle accurate level reference does not have such issue, as it emulate each time cycle of design. To keep the transaction level feature, a good work around is that reference model emulate several important time stages of stimulus's processing instead of each time cycle.

2.1 EMULATE STIMULUS' PROCESSING AT STIMULUS' START AND END.

Cycle accurate reference model can totally emulate every time step of stimulus' processing. However, it is not what we want, as the effort is almost like to re-write a RTL design and such model is quite difficult to maintain when real design suffers a little change. In fact, what the reference needs is just to emulate some important time points of stimulus' processing. It is not necessary to care about every state of stimulus' processing at each cycle.

Two of the most important timing points in stimulus' processing are stimulus' start points and end points. At beginning reference model is in known state, we set it verified state VS. Once a stimulus is injected to reference model, we can look the injection operation as the start point of the stimulus and put it into a timing uncertain stimulus set, we mark it as U_SET. This uncertain

stimulus set is bind to VS state. We mark this set it as U_SETVS. In addition, with simulation going on, VS state may evolve into several possible states due to stimulus competition scenarios. We mark the possible state as PS and manage all uncertain stimuli in an independent set, U_SET_ALL. U_SET_ALL include all stimuli, which are bind to different PS. We set all new added stimuli in U_SET_ALL as processing state for later use. Figure 3 shows the flow chart of managing injected stimulus by U_SETs (Different U_SET is bind to different PS or VS) and U_SET_ALL. To make it a more common solution, we assume reference model's starting state is in several possible states (PS). Each possible state (PS_i) has a U_SET bonded on it and marked as U_SET_{PS_i}.

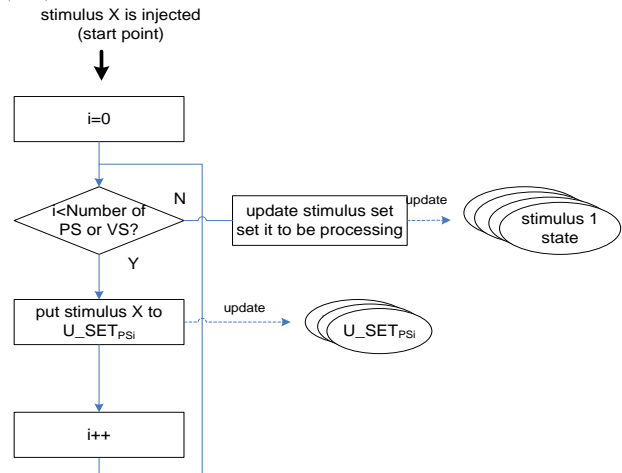


FIGURE 3 Manage input stimulus with U_SET_{PS_i} and U_SET_ALL

When an event of design is monitored by TB, we make evaluation to see if any stimulus end point is relative to this event. A good case is we are sure some stimuli are definitely relative to this event. We mark these stimuli as processed in U_SET_ALL. However, a more complicated case is we just know some stimuli are possibly relative to this event. We will handle this complicated case directly. Reference model should try to process the U_SETVS (or each U_SET_{PS_i}). The process step is as followed.

Step 1: figure out all possible sequences combination in U_SETVS (or each U_SET_{PS_i})

Some combinations maybe illegal and should be excluded: for example, if two stimuli come from same interface, one must be injected later than the other. The sequence combination can be managed by a stimulus queue and the queue should be processed with its bind state VS as reference model's initial state. If reference model is in several possible states, each U_SET, which is bind to its possible state (PS) can make a bunch of queues. We managed this bunch of queues with a queue set marked as queue_set_{PS_i}. For easy use of next step, we put all these queues into a big queue set (queue_set_{all}). Each queue will be attached with a reference model's initial state PS_{init,i}. PS_{init,i} is equal to queue's bind state (PS_i or VS) at start. We also attach J_{veri-end} and J_{min-end} to each

queue and set the queue as processing state for later use. The meaning of $J_{veri-end}$ and $J_{min-end}$ will be explained later. Figure 4 shows the flow chart of this step.

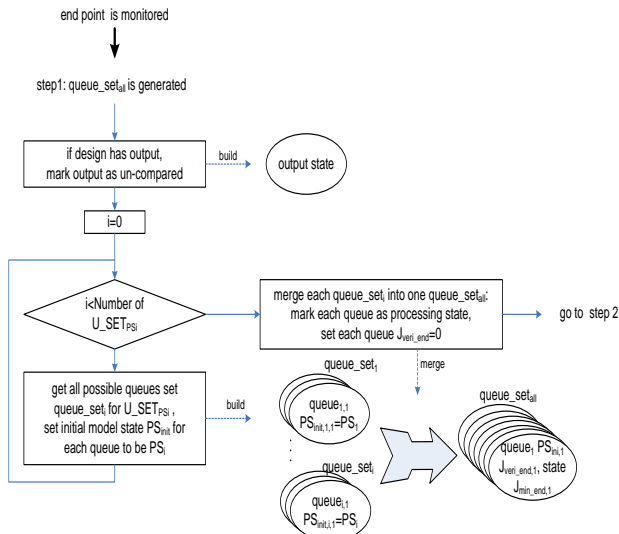


FIGURE 4 Generate the queues for all possible stimulus sequence combination

Step 2: Process all queues in queue_setall with their attached state as reference model's initial state.

For queue i , we process these stimuli one by one and from start to end. As a stimulus may influence design's output, we can compare reference model's output with design's to make sure if this stimulus's end point happens or has been processed by design. A variable $J_{veri-end,i}$ is used to record the position of latest stimulus whose end point is verified to happen. In addition, we need a variable $TEMP_PS$ to record model's state for next stimulus's processing. It is updated after a stimulus's processing. According to comparison result, we get two routines depending on the comparing result: If reference model's output match with part or whole of design's output, evaluate this stimulus to see if current reference model's output is influenced by it. If the influence exists, update $J_{veri-end,i}$ to the stimulus's position and update reference model's state $PS_{init,i}$ to $TEMP_PS$. We also update this stimulus's state as processed in U_SET_ALL for later use. Then continue the next stimulus processing. One thing should be noted that to check if output is influenced by current stimulus is a case-by-case problem and should be carefully figured out during implementation.

If reference model's output does not match with any of design's output, stop this queue's processing and start next queue's processing.

During one queue's processing, we can check if design's output is all compared off. If it is, that will mean design's behaviour is right. If not and all queues are processed, that means some error happens and we need to check design or model in further. Figure 5 shows this step's flow chart.

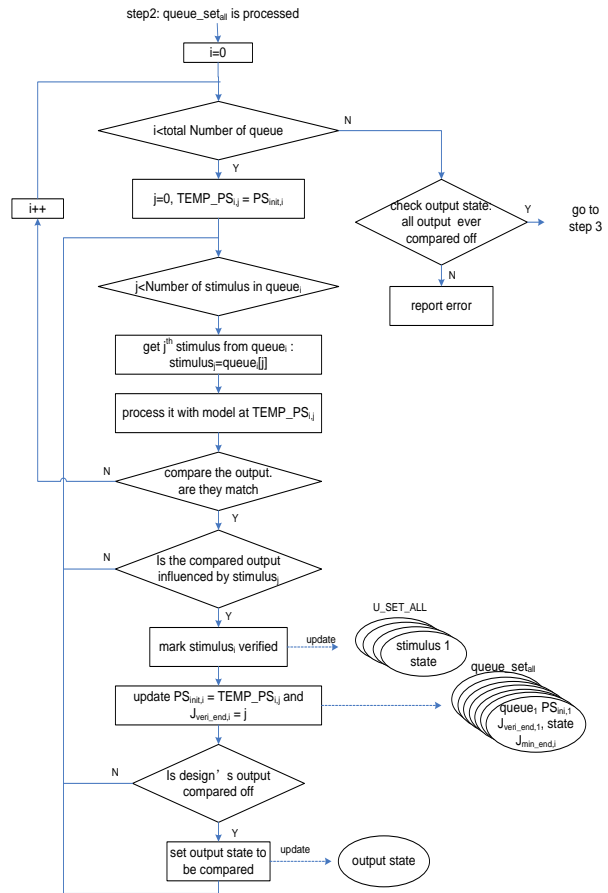


FIGURE 5 Process all queues in queue_setall to queue's end

Step 3: Process all queues in queue_setall to real $J_{min-end}$.

In previous step, we can figure out which stimulus's end point happens and already mark it as processed state in U_SET_ALL . Now we should process each queue again and update their attached state forward until latest stimulus which is marked as processed in this queue. For queue i , the process can be start with $J_{veri-end,i} + 1$ and initialized state $PS_{init,i}$, as $PS_{init,i}$ has been updated with $J_{veri-end}$ stimulus's processing in previous step. According to the state of each stimulus in whole uncertain stimulus set U_SET_ALL , we can find out the latest stimulus whose end point happens in queue i . We mark its position as $J_{min-end,i}$. It is a similar flow as step 2: after one stimulus is processed, by comparing the output with design's, we get two routines depending on the comparing result.

If reference model's output match with part or whole of design's output, Update $J_{veri-end,i}$ to the stimulus position and update reference model's state $PS_{init,i}$. Then continue the next stimulus processing.

If reference model's output doesn't match with any of design's output, stop this queue's processing and mark this queue as discarded. And then start next queue's processing.

After all queues are processed, check the state of all queues. If they are all in discarded state, some error

happens and we need to check design or model in further. The flow chart is showed in Figure 6:

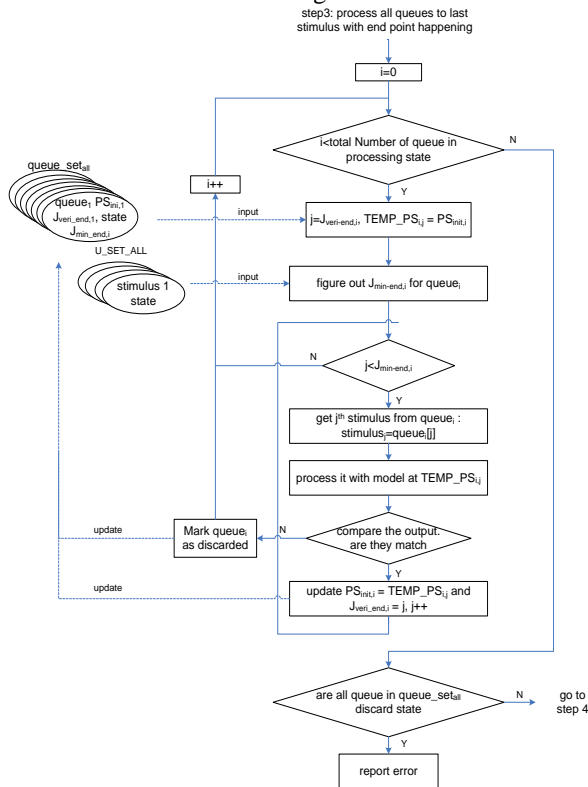


FIGURE 6 Process all queues in queue_set_all to real $J_{min-end}$

Step 4: merge the left queues.

After step 3, for queue_i, reference model's state is updated to a new PS_{init,i} and there are some stimuli (index from J_{veri-end,i} + 1 to end) are left as new uncertain stimulus set which is bind to PS_{init,i}, U_SET_{PSinit,i}. We can compare queue_i with all other queues. If some of them are equal (their left stimuli are equal and binding state PS_{init,i} are equal too), they can be merge to one U_SET_{PSi}. Then we get several new possible states, PS_i and new stimulus set U_SET, which is bind to it and is ready for next around processing.

2.2 AN APPLICATION FOR THE EXAMPLE DESIGN IN CHAPTER 2

Following this method, it is not hard to figure out a solution for the example mentioned chapter 2. Assuming current verified state VS is buffer full, packet A is coming and then a flush command B follows shortly after. The uncertain stimulus set U_SET_{VS} will have packet A and command B. So does U_SET_{ALL}. Due to flush command B, design will output packet payload in buffer and mark flush command B's end point. Therefore, we can set flush command B as processed. To make the competition stimulus scenario happen easily, we assume buffer can store two packets' payload.

Two possible sequences combination in U_SET_{VS} are got: flush command B -> packet A and packet A -> flush command B. they build up queue_set_{all} too.

For queue 1, after flush command B is executed, reference model will output data in buffer. TEMP_PS is buffer empty. PS_{init,1} is updated to TEMP_PS, buffer empty state. Then for packet A, it will be saved according to TEMP_PS. However, design should not have packet A's payload as output. So Packet A is still in processing state and J_{veri-end,1} keeps to be 1. For queue 2, packet A is discarded and reference model will output data in buffer after flush command B. Flush command B can be verified. PS_{init,2} is updated to buffer empty state, J_{veri-end,2} is set to 2.

By checking U_SET_{ALL}, we can figure out J_{min-end,1} is 1 and J_{min-end,2} is 2. So step 3 can be ignored. Now two queue's state is as followed: buffer empty and packet A is left in queue; buffer empty and no stimuli. They cannot be merged. Therefore, we get U_SET_{PS1} and U_SET_{PS2}.

Then come packet C, packet D and flush command E, which may compete with packet D. The new packet input is also an event of end point of previous packet. Therefore, packet A and packet C can be marked as processed. Here we skip the end event of packet A and packet C to ignore the procedure we do not care. For U_SET_{PS1} we have two possible queues: packet A->packet C ->packet D->flush command E and packet A->packet C -> flush command E -> packet D. For U_SET_{PS2} we have two possible queues too: packet C ->packet D->flush command E and packet C -> flush command E -> packet D. Now we can merge them into one big queue_set.

Queue 1: packet A-> packet C ->packet D->flush command E, PS_{init,1} is buffer empty.

Queue 2: packet A-> packet C -> flush command E -> packet D, PS_{init,2} is buffer empty.

Queue 3: packet C -> packet D -> flush command E, PS_{init,3} is buffer empty.

Queue 4: packet C -> flush command E -> packet D, PS_{init,4} is buffer empty.

For queue 1, packet A, packet C will be saved; Packet D is discarded due to buffer is full again; Packet A and packet C will be flush out due to flush command E. Now we can get sure state about packet A by comparison result with design's output. If two results match, packet A and packet C are set to processed state. Queue 1 is updated to the state that buffer is empty and no stimulus is left in queue. If results don't match, queue 1 will not be updated.

For queue 2, packet A, packet C will be saved and then be flushed out. Packet D is saved. By comparison result with design's output. If two results match, packet A and packet C are set to processed state. Queue 2 is updated to buffer empty and packet D is left in queue. If results do not match queue 2 will not be updated.

For queue 3, packet C and packet D are saved and packet C will be flush out.

For queue 4, packet C and packet D are saved and flushed out. As the output is different, we can make sure if packet D is flush out or not.

If packet A is not discarded in previous stimulus competition scenario, queue 3 and 4 will be marked as discarded state in step 3. Queue 1 and queue 2's states are quite like the states of two queues in previous competition scenario. If packet A is discarded, queue 1 and queue 2 will be marked as discarded state in first stimulus's processing and comparing. Queue 3 and queue 4 will make a duel based packet D's comparison. So only one of them survives and bring reference model to a determined state.

3 The consider factor in reference model's implementation

Sometimes there are too many stimuli in the U_SET, which will make too many PS states. Based on these new PS states, a new big U_SET may be got again. After processing these new U_SET, more second stage of PSs may be got. Chapter 3.2 has shown us an example, 4 queues and two PSs are got in second stimulus competition scenario. To record and maintain these PSs and U_SET is a complicated job. So better to do some optimization or trade off during reference model's implementation in a verification system.

3.1 REDUCE THE AMBIGUOUS TIME

For stimulus competition scenarios which may generate different result due to different time delay of stimulus' processing, time between start point and end point of the stimulus can be regarded as an ambiguous time. We do not know design behaviour definitely. If the ambiguous time is reduced, the number of stimuli in U_SET will be reduced and the number final possible state PS after all possible queues are processed will be reduced too. To do this, we need to mark the stimulus start point as later as possible and mark stimulus end point as early as possible. For example, set stimulus start point when the stimulus are totally injected to DUT if we are sure that design output can only be affected after the whole stimulus are totally taken in by design.

If internal signal of design is available to verification engineer, checking the internal signal of design is another good method. Although internal signal can be changed, some important signals are usually preserved if functionality of design is not modified a lot. Moreover checking internal signal means we can let reference model sync to design's timing step and can reduce ambiguous time to 0. This method is a trade-off. If too many internal signals are monitored, to maintain these signals will be another burden for testbench. A worst case is sometimes RTL is encrypted if verification system is developed by third part agent.

3.2 CONSTRAIN THE STIMULUS'S GENERATION

The PS or VS of reference model can be a good feedback to generator. When competition stimulus scenario

happens and several possible results are got. By checking the state of reference model, the generator can be forced to generate stimulus, which can bring definitely determined state to reference model. For example in chapter 3.2, if generator finds that reference model is not sure about buffer's state as packet A may be discarded or saved, another flush command can be injected. Then by comparing flush output, packet A's state will be determined quite soon. However, with such constraint, some scenarios cannot be produced. Therefore, verification engineers should judge if lost scenarios are important for the design's function verification.

Another advantage of constraining stimulus's generation based on reference model's state is the expected corner case is easier to hit than normal randomized generation. For the example design given in chapter 2, buffer full can be a corner case. By checking buffer state in reference model, we can constraint to generate more packets and less flush command when buffer is nearly full. Then buffer full condition can be easily achieved.

3.3 SUBTRACT THE EVENT FOR STIMULUS'S START POINT AND END POINT

Subtract the right event for stimulus's start point and end point is another important factor for the method presented in this paper. In fact, from a broad concept, polling design's internal signal is a way to subtract the event of stimulus's start point and end point. However, by analysing design behaviour, some event can be subtracted just based on the input and output.

You will find that solution in chapter 3 will not work if we modify the design's behaviour like this: Interface C must be in a stable speed if data flush is ongoing. Due to buffer and speed of interface A, C are well defined, when the buffer is only flushed a little, available byte cannot afford the incoming packet, the packet will be discarded; when the buffer is flushed too much, to flushing out incoming packet may face the risk that left byte in buffer is flushed out but incoming packet is not fully received. Under such condition, design may not keep a stable flushing speed for interface C. So incoming packet will be just saved; when the buffer is flushed neither so less nor so much, the incoming packet can be flushed out.

The answer for such design change is not hard: if we subtract a stimulus like "enough byte is flushed out for incoming packet" which is after the flush command and adapt it to the solution, the problem can be solved again.

4 Summary

The method to implement reference model presented by this paper is to find all possible states and results during ambiguous time and figure out the final state and result after comparison with design's output. To do this, reference model will record all stimuli whose end points are not coming. By process all possible sequence for

these stimuli, design's possible state and output will be figured out. From 4.2 we also find an extra advantage to record these states, which help with the test to cover the specified corner case easily. This is another back up to implement reference model by using this method.

References

- [1] Bergeron J 2006 *Writing testbenches using system Verilog* Springer.
- [2] Fitzpatrick T, Salz A, Rich D, Sutherland S 2006 *System Verilog for Verification* Springer
- [3] Rosenberg S, Meade K A 2010 A practical guide to Adopting the universal verification methodology (UVM). Cadence Design Systems.
- [4] Haque F, Michelson J 2001 *Art of Verification with VERA* Verification Central
- [5] Jindal R, Jain K 2003 Verification of transaction-level SystemC models using RTL testbenches. Formal Methods and Models for Co-Design, 2003 MEMOCODE'03. Proceedings. First ACM and

Acknowledgments

Our work is supported by the National nature science foundation of China (NO.61103161) and the Program for New Century Excellent Talents in University (NO. NCET-12-0579).

IEEE International Conference on, IEEE (Mont Saint-Michel, France, 24-26 June) 199-203

- [6] Rowen C 2002 Reducing SoC simulation and development time *Computer* 35(12) 29-34
- [7] Clouard A, Mastrococco G, Carbognani F, Perrin A, Ghenassia F 2002 Towards bridging the precision gap between SoC transactional and cycle-accurate levels *Proc. Design, Automation and Test in Europe (DATE, Paris, France, 4-8 March, 2002)*
- [8] Carbognani F, Lennard C K, Ip C N, Cochrane A, Bates P 2003 Qualifying precision of abstract systemc models using the systemc verification standard. Design, Automation and Test in Europe Conference and Exhibition, 2003, IEEE (Munich, Germany, 3-7 March) 88-94

Author



Lirong Qiu, born on August 28, 1978, China

Current positions, grades: full professor of computer sciences at Information Engineering Department, Minzu University of China
University studies: M.Sc. in Computer Sciences (2004) and PhD in Information Sciences (2007) from Chinese Academy of Science.
Scientific interests: different aspects of natural language processing, artificial intelligence and distributed systems.

Application of fuzzy comprehensive evaluation in weapon equipment systems

Rong-chun Wu^{1, 2}, Feng-li Zhang¹, Jin-bang Zhang², Qian He^{2*}

¹School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

²Second B. Police Academy, Chengdu, 610213, China

Received 1 March 2014, www.tsi.lv

Abstract

Analysis and evaluation of the operational effectiveness of weapon equipment operational systems has always been a complex problem, a study of its evaluation technology is of great significance. Task oriented operation, this paper discusses equipment operational system dynamic integration needs, and discusses the steps and comprehensive performance evaluation method of weapon equipment operational system, made a fuzzy comprehensive evaluation method to better adapt to comprehensive performance evaluation of weapon equipment operational system. According to the characteristics of system evaluation factors, and gives an indicator of quantitative methods established based on analytic hierarchy process and correlation analysis of comprehensive performance evaluation model, based on three types of weapon equipment operational system data, for example, proves the validity of the method.

Keywords: Weapon Equipment Systems, Comprehensive Performance, Combat Effectiveness, Multi-level Fuzzy comprehensive evaluation

1 Introduction

With the weapons and equipment have been modernized and weapons and equipment the ability to play in the combat system, military experts alone, intuition and experience to evaluate, it is impossible to meet the requirements of high-tech war under the current conditions, fast, accurate and effective equipment operation evaluation methods have become weapons and equipment development is bound to ask. Evaluation of weapon equipment operational capability is the planning, development, equipment and weaponry deployment and operational application of important links. Played by weapons and equipment in the course of actual combat capability, mainly using operational effectiveness is measured by the weapons and equipment of combat missions performed by. Measure of the weapon equipment operational effectiveness in the combat system is one of the weapons and equipment of the most important assessment parameters; it can reflect the essential characteristics of the weapons and equipment in combat system. In the current world, military equipment operational effectiveness evaluation has become a research and development facility equipped with a "hot" topic. The operational effectiveness evaluation of equipment, equipment, combat simulation, simulation technology has played a key role in the application of, but as the battlefield situation information under the condition of information needs, simulation, simulation applications there are some new issues that cannot be resolved, For example, in the "linear features",

"dimensions of disaster" and "complexity of disaster" features such as handling performance in particular.

This article in the previous study of weapon equipment operational system operational effectiveness based on the analysis, considering the issue and proceed from the characteristics of weapon equipment operational effectiveness, presented to the operational effectiveness of weapon equipment operational system analysis of comprehensive evaluation method using fuzzy comprehensive evaluation method to build weapons and equipment effectiveness evaluation model and evaluation system of weapon equipment operational system effectiveness is given.

2 Operational effectiveness evaluation of weapon equipment operational system processes

Weaponry in combat system performance evaluation, not only demand for weapons and equipment appeared a clear understanding of the root of the problem, also need clear evaluation of data sources to determine what assessment methods are used, and map out the equipment operational effectiveness evaluation process to guide the evaluation studies. Literature [1] gives the typical evaluation of weapon equipment operational effectiveness evaluation processes (for example, as shown in Figure 1).

*Corresponding author e-mail: 2629000765@qq.com

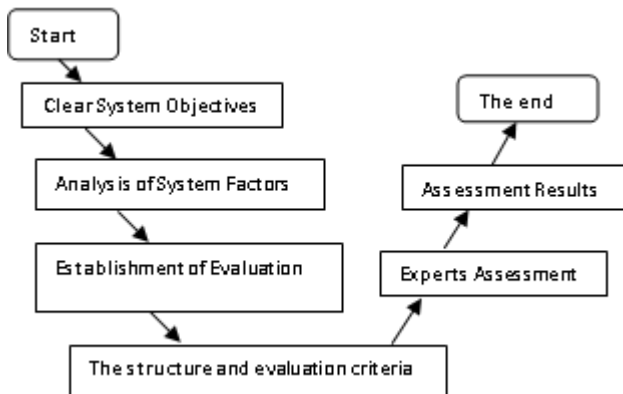


FIGURE 1 The evaluation of weapon equipment operational effectiveness processes

The above (Figure 1) evaluation processes in the area of evaluation of weapon equipment operational effectiveness is widely used, but it also has its disadvantages:

- 1) The lack of feedback loops in the middle of the evaluation process; it is a one-way assessment process. Satisfaction evaluation system needs to be able to pass a reasonable assessment came over and over again, and the process is repeated evaluation of the loop can be adjusted; one evaluation process is unable to meet this requirement.
- 2) The fixed evaluation of data sources. Conventional evaluation methods, mainly fixed data source select static analysis results after processing. However in the actual equipment operational evaluation in the course of its data sources are constantly changing. So fixed evaluation of data sources for in-depth evaluation studies, limitations are made, it is difficult to find variation occurring in the evaluation process.

3 Operational effectiveness evaluation model of weapon equipment operational system

3.1 THE OPERATIONAL EFFECTIVENESS EVALUATION OF KEY LINKS

The key elements of weapon equipment operational effectiveness evaluation system is mainly reflected in: evaluation model of the system of indicators, evaluation parameters determining the calculation of indicators, evaluation parameters and performance evaluations. Evaluation of operational effectiveness for weapon equipment operational system currently, probability models, the classic method of ADC, SCA based on Cybernetics and its index model method [2]. Evaluation modelling is the key to the evaluation process, as a Department of weapon equipment operational effectiveness evaluation of programmed implementation and evaluation parameters of calculation on the assumption; it maps the evaluation for effectiveness

evaluation model, effective formative assessment solution scenarios.

The application of fuzzy comprehensive evaluation in the weapon equipment operational system focuses on how to determine the weaponry of scientific rationality and the weight of evaluating indicators in combat system. Method for determining the weights of the evaluation indexes for three main categories [3]:

- 1) *Subjective weighting method.* Mainly, expert investigation of binomial coefficients, APH analytic [4] and the least square method.
- 2) *The weight method.* Mainly includes: multi-objective programming method, entropy method, component analysis, etc.
- 3) *The method of combination weighting.* Combination weighting method is to use a combination of subjective and objective weighting method uses a weighting method, perform two weighting method advantages. Equipment operational review process, we use fuzzy consistent matrix to determine weight, avoiding the APH AHP analysis on the adjustment problems of inconsistency of judgment matrix.

3.2 FUZZY COMPREHENSIVE EVALUATION METHOD

- 1) *Target set.* The target set for $u = (u_1, u_2, \dots, u_n)$.
- 2) *Evaluation set.* The evaluation set for $v = (v_1, v_2, \dots, v_n)$.
- 3) *Fuzzy comprehensive evaluation model* [5, 6]. Due to the a-level fuzzy comprehensive evaluation:

$$B = (b_1, b_2, \dots, b_m) = (a_1, a_2, \dots, a_n) \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nm} \end{pmatrix}. \quad (1)$$

$$0 \leq r_{ij} \leq 1$$

The comprehensive evaluation of the first, after judge sets is going to judge the result as a single factor, two-level fuzzy comprehensive evaluation, whose formula is:

$$C = A \cdot B = \begin{bmatrix} A_1 \cdot R_1 \\ A_2 \cdot R_2 \\ \vdots \\ A_n \cdot R_n \end{bmatrix} = A \cdot \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} = A \cdot (b_{ij})_{n \times m}. \quad (2)$$

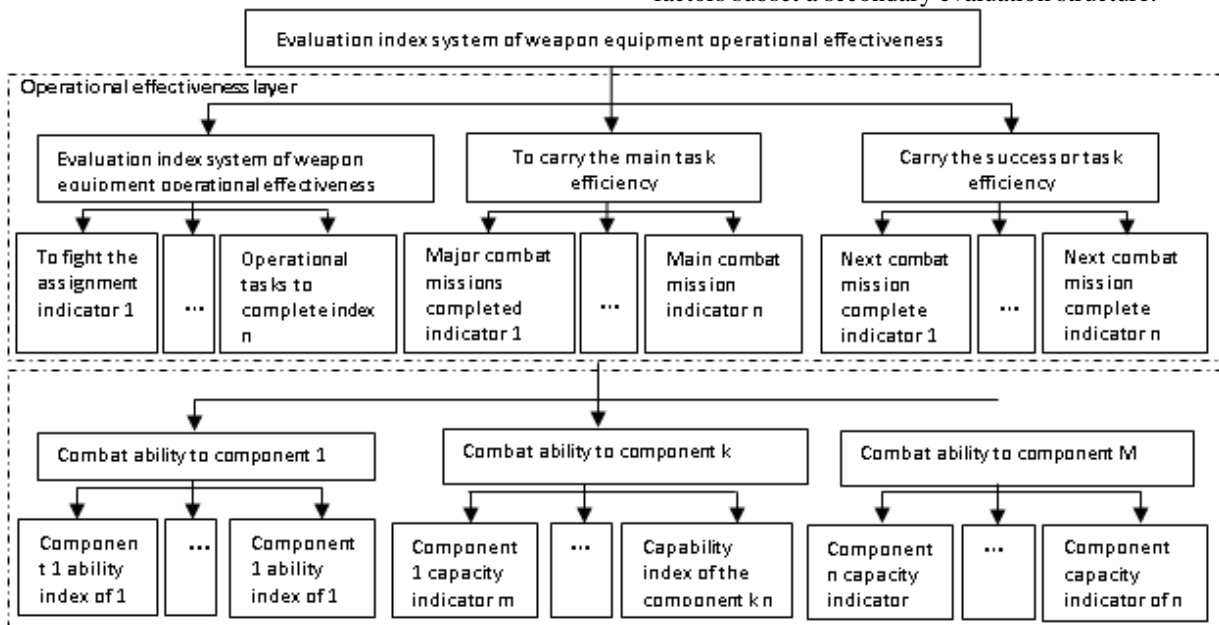
3.3 MODEL SYSTEM OF WEAPON EQUIPMENT OPERATIONAL EFFECTIVENESS

The equipment operation system effectiveness assessment is mainly on programmers of different weaponry systems and systems analysis, comparison; you

need to identify measure indicators. Indicators and combat effectiveness evaluation of weapon equipment operational tasks are closely related and, therefore, have appropriate characteristics, such as indicators to reflect the end of Figure thing, Figure thing for different purposes, their efficiency evaluation indicators are also different; indicators to reflect the randomness of combat, using a probabilistic nature of digital character to represent; indicators to reflect the complexity of the battle, with gradation and diversity. Armed combat system effectiveness evaluation by using the method of fuzzy comprehensive evaluation for effectiveness evaluation. First of all, to levels of operational efficiency evaluation systems of all kinds of weapons into fuzzy subset defines evaluation levels:

$$V = \{v_1, v_2, v_3, v_4, v_5\}. \tag{3}$$

Equipment operation system effectiveness assessment is divided into five levels: excellent (v_1), Good (v_2), Middle (v_3), General (v_4) and Poor (v_5). Corresponding values are as follows: 0.80 above and 0.79~0.60, 0.59~0.50, 0.49~0.40, and under 0.40. For different subsets of assessment, the meanings of the elements in the collection are not the same. To establish the assessment factors system structure diagram as shown in Figure 2. Evaluation factors is a subordinate relations associated class hierarchy, give class hierarchy is multistage evaluation factors system, and the top evaluation target. Here will be summarized as evaluation factors subset a secondary evaluation structure.



Level measurement set is:

$$U_1 = \{u_1, u_2, u_3\}, \tag{4}$$

where u_1 task before you can carry effectiveness; u_2 to carry the main task effectiveness; u_3 is the carry the successor task effectiveness.

The second-level test set is a family of sets:

$$U_2 = \{u_{1m_1}, u_{2m_2}, u_{3m_3}\}, \tag{5}$$

where $U_{im} = \{u_{i1}, u_{i2}, \dots, u_{im}\}, i = 1, 2, 3,$
 $m_i = 1, 2, \dots, M_i.$

By determining the parameters of the evaluation factors weight at all levels, come to the evaluation factors weight fuzzy membership vector parameters.

Level evaluation factors weight fuzzy sets as follows:

$$A_1 = \{a_1, a_2, a_3\}. \tag{6}$$

Secondary fuzzy evaluation factors weight set is a family of sets:

$$A_2 = \{A_{1m_1}, A_{2m_2}, A_{3m_3}\}, \tag{7}$$

where

$$A_{im} = \{a_{i1}, a_{i2}, \dots, a_{im}, a_{iM_i}\}, i = 1, 2, 3, \tag{8}$$

$$m_i = 1, 2, \dots, M_i.$$

a_{im} of A_{im} is the evaluation factors in the fuzzy set membership.

According to the rating evaluation factors, first establish a mapping of each grade from U to V , establish distribution fuzzy to determine fuzzy relation matrix. Describe the evaluation factor in accurate mathematical

models through formulas to calculate values, later transformed into fuzzy fields within the membership. Using membership function finds the fuzzy relation matrix evaluation factor fuzzy relation matrix elements in r_{ij} .

The corresponding evaluation grade $v_j = \{j = 1, 2, 3, \dots\}$. The degree of membership $a_{ij}(u_i)$ where $u_i = \{u = 1, 2, 3, \dots\}$.

Based classification evaluation factors assessment for each grade of membership function is normal, its fuzzy distributed [7, 8]:

$$a_{ij}(u_i) = \exp \left[- \left(\frac{u_i - m_{ij}}{\delta_{ij}} \right)^2 \right] \tag{9}$$

Among them $a_{ij}(u_i)$, m_{ij} and δ_{ij} , respectively, for the first i evaluation factors on the first u_i , j , v_j the membership degree of evaluation grades and their statistical values of mean and variance

$$R_i = \begin{bmatrix} r_{i_{1 \times 1}} & r_{i_{2 \times 1}} & \dots & r_{i_{4 \times 5}} \\ r_{i_{2 \times 1}} & r_{i_{2 \times 2}} & \dots & r_{i_{2 \times 5}} \\ \vdots & \vdots & \vdots & \vdots \\ r_{i_{m \times 1}} & r_{i_{m \times 2}} & \dots & r_{i_{m \times 5}} \end{bmatrix} \tag{10}$$

For model calculations, by determination of level measurement: $B = A \circ R = \{b_1, b_2, b_3, b_4, b_5\}$.

This operation model fuzzy operator for $M(\bullet, v)$:

$$B = A \circ R = \{b_1, b_2, b_3, b_4, b_5\}$$

Operations for the above levels in turn, calculated level of evaluation findings: $B = A \circ R = A_1 \circ (A_2 \circ A_1)$.

According to the principle of maximum membership degree judgment, their evaluation results are b corresponds to the maximum membership degree in evaluation of language [9].

4 Examples of application

Existing A, B, C three types of weapons and equipment combat system, according to the requirements of the soldiers of equipment capability, intends to select fire damage ability, field survival ability, mobility; the adaptability of battlefield four aspects of the comprehensive ability inspects the comprehensive performance of the combat system equipment.

TABLE 1 the performance of the three types of weapons and equipment combat system technical index

Capacity indicator	Performance indicators	Type series		
		Model A	Model B	Model C
Fire damage	Ammunition power	Excellent (0.75)	Excellent (0.75)	Good (0.625)
	Range	Excellent (0.75)	Excellent (0.75)	Excellent (0.75)
	Precision	Medium (0.50)	Medium (0.50)	Good (0.625)
	Measure aiming at the response time	Medium (0.50)	Excellent (0.75)	Good (0.625)
	Fire duration	Medium (0.50)	Medium (0.50)	Medium (0.50)
Battlefield survivability	Camouflage ability	Good (0.625)	Medium (0.50)	Medium (0.50)
	Protection capability	Good (0.625)	Excellent (0.75)	Good (0.625)
	Risk perception	Medium (0.50)	Good (0.625)	Medium (0.50)
	Work time	Good (0.625)	Medium (0.50)	Good (0.625)
Mobility	total weight	Good (0.625)	Excellent (0.75)	Medium (0.50)
	Overall dimensions	Medium (0.50)	Medium (0.50)	Medium (0.50)
	Time to prepare	Good (0.625)	Medium (0.50)	Good (0.625)
Battlefield adaptability	reliability	Poor (0.25)	Medium (0.50)	Medium (0.50)
	availability	Medium (0.50)	Medium (0.50)	Medium (0.50)
	Service features	Medium (0.50)	Good (0.625)	Poor (0.25)

The influence factors of the efficiency evaluation set up $V = \{v_1, v_2, v_3, v_4, v_5\}$, including v_1, v_2, v_3, v_4, v_5 respectively corresponding to the excellent, good, medium and poor. The values above 1.0, 0.75, 0.625, 0.75 and 0.625 below, (ex: as shown in Table 1).

Using analytic methods, determining each factor index on the upper level index weighting is as follows:

$$w = (0.428, 0.160, 0.128, 0.074),$$

$$w_1 = (0.458, 0.041, 0.162, 0.256),$$

$$w_2 = (0.053, 0.548, 0.102, 0.297),$$

$$w_3 = (0.731, 0.081, 0.188),$$

$$w_4 = (0.637, 0.258, 0.105).$$

4.1 INTEGRATED PERFORMANCE INDICATOR LEVEL OF EVALUATION

Table 1 performance of weapon equipment operational systems in fire damage capability properties matrix is:

$$R_1 = \begin{bmatrix} 0.75 & 0.75 & 0.625 \\ 0.75 & 0.75 & 0.75 \\ 0.50 & 0.50 & 0.625 \\ 0.50 & 0.75 & 0.625 \\ 0.50 & 0.50 & 0.50 \end{bmatrix}.$$

A planning is: $R_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}.$

The optimal vector $G_1 = (1, 0, 1, 1, 0)$, $B_1 = (0, 0, 0, 0, 0)$, known weight vector for $w_1 = (0.458, 0.041, 0.162, 0.256)$ can be calculated for each vector-valued, generation into the evaluation model, the evaluation results will be achieved.

$$u_1 = \{0.6490, 0.8349, 0.2247\}.$$

Similarly, it can be obtained:

$$u_2 = \{0.2462, 0.7538, 0.1752\},$$

$$u_3 = \{0.8228, 0.8795, 0.1205\},$$

$$u_4 = \{0.1008, 0.9208, 0.8396\}.$$

4.2 COMPREHENSIVE PERFORMANCE INDEX OF THE SECONDARY EVALUATION

$$R_1 = \begin{bmatrix} 0.6490 & 0.8349 & 0.2247 \\ 0.2462 & 0.7538 & 0.1752 \\ 0.8228 & 0.8795 & 0.1205 \\ 0.1008 & 0.9208 & 0.8396 \end{bmatrix}.$$

The optimal vector $G = (0.8349, 0.7538, 0.8795, 0.9208)$, $B = (0.2247, 0.1752, 0.1205, 0.1008)$, known weight vector for $w = (0.428, 0.160, 0.128, 0.074)$ can be calculated for each vector-valued, generation into the evaluation model,

References

[1] Przemieniecki J S 2009 *Fuzzy Sets and System* **78**(5)113-52
 [2] Liang G S 2011 *European Journal of Operational Research* **112**(4) 36-9
 [3] C4ISR Architecture Working Group. *C4ISR Architecture Working Group Final Report* 1998 USA:Department of Defense
 [4] Brass P 2007 *ACM Trans. Sen. Netw.* **3**(2)
 [5] Chen Hua-you, Chen Qi-ming, Wang Hui 2008 Uncertain aggregation method on group decision making based on relative entropy *Journal of Anhui University (Natural Sciences)* 2008-01 (in Chinese)
 [6] Zhou Yuan 2012 Comprehensive evaluation model of public transit network based on Non-negative matrix factorization *Electronic Design Engineering* 2012-11 (in Chinese)

the evaluation results will be achieved:
 $u = \{0.1670, 0.9783, 0.0458\}.$

4.3 EXPERIMENTAL EVALUATION OF THE RESULTS

Through the above evaluation results of the comparison, from the size of the comprehensive performance, A model > B model > C model, the in conformity with the actual situation is. Compared with other comprehensive performance evaluation method, the result is reasonable, more practicable methods.



5 Conclusions

Weapons in combat role in the process of weapon equipment can be used in the process of operations in ability to measure stipulated task. Combat system effectiveness evaluation is one of the most important evaluation indexes of weapon equipment system; it can reflect the essence of weapon system. Using fuzzy comprehensive evaluation method, through the establishment of weapons and equipment combat system evaluation model, can be found by the experimental results on the efficiency of weapon equipment operational system qualitative quantitative research. Give full play to command and combat effectiveness, to make the equipment system effectiveness is stronger than the enemy.

Acknowledgments

This work was supported by the National Science Foundation (Grant No. 61133016), and the National High Technology Joint Research Program of China (863 Program, Grant No. 2011AA010706), thank for the help.

[7] Sun Xiao-ling, Wang Ning 2013 Interval Valued Fuzzy Evaluation of the Comprehensive Quality of College Students *Journal of Hefei Normal University* 2013-03 (in Chinese)
 [8] Cai Lei 2006 *Fuzzy Synthetic Evaluation in the Establishment of Basic Science Subject[A]* [C] (in Chinese)
 [9] Liu Buquan, Wang Huaimin, Yao Yiping 2005 Data Consistency in a Large-Scale Runtime Infrastructure *Proceedings of the 2005 Winter Simulation Conference* Orlando FL USA
 [10] Tillal Eldabi, Man Wai Lee, Ray J P 2004 Examining the Feasibility of Constructing Simulation Models Using the Web-Based 'Grab-and- Glue' Framework *Proceedings of the 2004 Winter Simulation Conference* Washington DC USA
 [11] Robert L W 2000 AUV Commercialization-Who's Leading the Pack *Proceedings of Oceans 2000 MTS/IEEE conference and Exhibition* Providence Rhode Island 391-5

Authors	
	<p>Wu Rong-chun, born on December 6, 1976</p> <p>Current position, grades: doctoral candidates, works at Police Academy, lecturer Scientific interest: Information fusion, command automation</p>
	<p>Zhang Feng-li, born April 6, 1963</p> <p>Current position, grades: Doctor, Professor, a tutor for doctors, works at University of Electronic Science and Technology School of Computer Scientific interest: the database, the syncretism of ambulation data</p>

Expert system based on fuzzy rules for maritime search and rescue

Yingjun Zhang*, Xuefeng Yang

Navigation College, Dalian Maritime University, Dalian, China Linghai Road #1, Dalian, China, 116026

Received 1 June 2014, www.tsi.lv

Abstract

Search and rescue (SAR) plan decides the result of SAR activity, and relates with the safety of life and property at sea. To improve the efficiency and standard of SAR, and give the SAR officer support to make better decisions, expert system (ES) is been researched by this paper, and the ES based on fuzzy rules for maritime SAR is proposed. Firstly, the structure of ES based on fuzzy rules for SAR is designed. Secondly, we have researched SAR knowledge acquisition and knowledge representation, chosen five ways to acquire SAR knowledge. At last, we designed the inference engine of ES for SAR, and introduced it in an example.

Keywords: fuzzy rules, expert system, supporting system, search and rescue

1 Introduction

With the economic development exchange of commodities among countries become more frequent, the amount of ship being engaged in transporting is increasing continuously, and the scale of maritime activities is more and more big, which make the task of maritime search and rescue (SAR) more heavy [1]. SAR is not only a difficult task, but international responsibility [2]. The safety of shipping has been improved by the advanced navigation technical. However there are complex and unpredictable situation at sea, and maritime accident still happens frequently. For example, in June 2013 "Asian Express" ship with Maldives nationality damaged and sunk after floated two days in heavy weather because of power lost in Arabian Sea. In 2011, general cargo vessel "Swanland" sunk in Irish Sea due to hull damage. In June 2008, passenger ship "Star Prince" sunk in waters of the central Philippines after encountered typhoon and more than 700 people lost their lives. In February 2006, passenger ship "Salamu 98" sunk after explosion, and more than 1000 people were killed. After maritime accident the life of passengers and crew is under threat, and SAR is very necessary. Maritime SAR is not only the method to guarantee the safety of life and property but the expression of international humanitarian.

SAR refers to the search and rescue action made by search and rescue force except the ship in distress after they acquire distress message, which contains search section and rescue section. Search means to determine the location of people in distress with the coordination of SAR coordination centre, and rescue indicates to save the people in distress, provide preliminary medical service and other necessary service for them, and move them to safe place [3]. In order to improve the efficiency of SAR, researchers in different country has done a lot of work

about SAR, and the emphasis of their work would be shown in the following:

(1) The search area. For example, reference [3] proposed an ensemble drift model of search objects based on stochastic particle simulation approach to determine the search area.

(2) Optimal model for the selection of search and rescue force at sea. For example, reference [4] constructs the optimal model for search force selection at sea, solves the model, and simulates the search situation.

(3) Supporting system for maritime SAR. Because the SAR problem is very complex, commanding officer cannot make perfect SAR plan according to his own experience. In response to this situation, decision supporting system based on cases [5] and expert system [6-10] were used in SAR; assist the officer to make SAR plan.

Expert system (ES) is one part of artificial intelligence, and is a computer program that solves the specialized problems at the level of a human expert [11]. Compared with human experts, the characteristic of ES is that it can study and explain the process of inference, adapt difficult environment, inference and deal with big data. If the ES is used, the efficiency of SAR would be improved. Nowadays, there are different kinds of ES, mainly contains ES based on rules, ES based on frame, and ES based on cases. Among them ES based on rules is widely used because of its simple development tool and successful example, so this paper researched the ES based on fuzzy rules for maritime search and rescue.

2 Search and Rescue Expert System

2.1 STRUCTURE OF SAR ES

The basic structure of ES based on rules was proposed

* *Corresponding author* e-mail: ecdis@sina.com

by Newell and Simon in Carnegie Mellon University, which contains five parts. They are respectively knowledge base, database, inference engine, explanatory equipment and user interface. On this basis, this paper designs the structure of ES based on fuzzy rule for maritime search and rescue, and its structure is shown in Figure 1. The system contains six parts, they are respectively SAR Knowledge base, SAR work memory, inference engine, explanatory program, knowledge acquisition, and user interface.

SAR knowledge base contains all knowledge about maritime search and rescue domain. In this system, SAR knowledge is represented with a lot of fuzzy rules. Each rule indicates a relationship, suggestion, instruction, tactic, or heuristic method, and has IF (LHS: Left-Hand-Side) THEN (RHS: Right-Hand-Side) structure. When the LHS is met, the rule is aroused, and the RHS is implemented. According to the function of SAR knowledge, we can divide them into maritime search knowledge and maritime rescue knowledge. The former is used to decide the location of people in distress, and search plan. The latter is used to save the people in distress after they are discovered. Therefore, the search knowledge is basic of SAR plan, and the rescue knowledge is the leader of rescue scene.

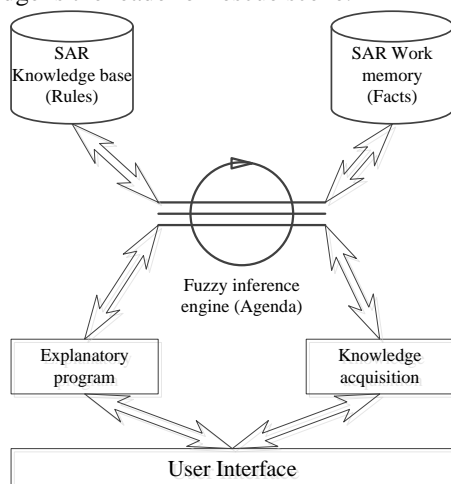


FIGURE 1 The Structure of ES based on fuzzy rules for SAR

SAR work memory contains a set of facts, which is used to match LHS part of knowledge base. These facts refer to the maritime SAR information, which has a good variety, and contains the information of ship in distress, SAR force, and meteorology and sea condition. Fuzzy inference engine's mission is to find the reasonable SAR plan according to SAR knowledge and SAR information. Additionally, explanatory program explains the inference process for user, which can be regarded as the basis of SAR plan. Knowledge acquisition is used to add new knowledge to system, and self study. User interface is the bridge between the user, human experts, knowledge engineers and ES.

2.2 DEVELOPMENT TEAM OF SAR ES

In order to introduce the development of ES, this part would introduce the development team of SAR ES. It contains five parts: SAR domain experts, knowledge engineers, programmers, manager and users, and the relationship among them are shown in Figure 2. Knowledge engineers are responsible for the work to design, construct, and test the ES, programmers are responsible for code, manager leads the process of development and coordinates other members, and users are mainly the officers who make SAR plan or commanding officer in rescue scene.

SAR domain experts should have the knowledge about shipping, maritime, aeronautics, medical, environment, petrochemical, ocean engineering, and meteorology. They are members of development team of ES, because they have experience in SAR. In SAR domain, experts work at maritime search and rescue centre or SAR scene. However, it is very difficult to let them represent their experience with rules. Additionally, the SAR knowledge provided by them is often fuzzy, for example, they may tell us that if the weather is bad, we should access the man overboard from the leeward. Therefore, we do not know when this rule should be activated due to "bad" is a fuzzy concept. Compared with traditional ES, such as EMYCIN, PROSPECTOR, the knowledge of ES for SAR is just experience, and variable. In addition, it can be represented by fuzzy rules.

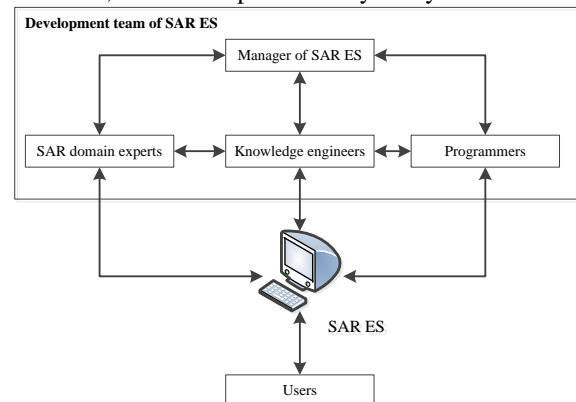


FIGURE 2 The development team of ES for SAR

3 Search and Rescue Knowledge

3.1 SAR KNOWLEDGE ACQUISITION

SAR knowledge acquisition is very important for the development of ES. During the process of development we carried different kinds of method to acquire knowledge, which contains International search and rescue manual, SAR expert interview, blind test, search and rescue cases, and other literatures. The source of SAR knowledge is shown in Figure 3.

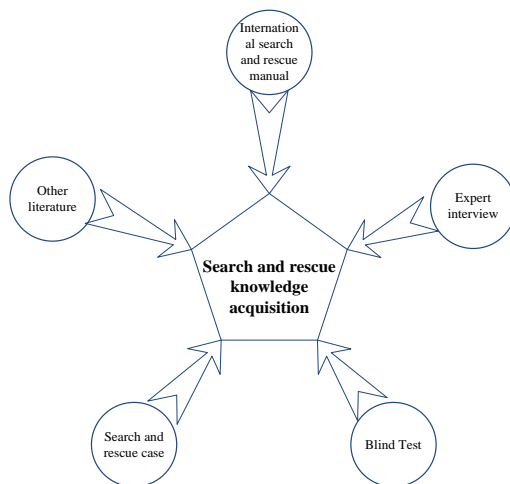


FIGURE 3 SAR knowledge acquisition

TABLE 1 SAR rules from international SAR manual

Rule 34	IF AND THEN	Ship does not arrive beyond its expiry date Ship does not report. The ship is in an unknown phase, inquire it first time.
Rule 35	IF OR THEN AND	There is no result after inquire it first time There is information to indicate the ship has been damaged. Alarm Inquire it once again by expanding the inquiry scope
Rule 36	IF OR THEN	There is still no result after second inquiry There is information to indicate the ship is in distress The ship is in distress

International SAR manual: international SAR manual mainly contains handbook of international aviation and maritime search and rescue, merchant ship search and rescue handbook and maritime search and rescue handbook. These handbooks contain much SAR knowledge, are crystal of SAR workers' wisdom. They provide us with organization method, search method, and rescue method. The SAR knowledge from these handbooks has authority. After analysed them, this paper generalized many SAR rules. There are some examples shown in Table 1.

Expert interview: expert interview can let us know what is necessary for SAR inference, regard them as node, and then we can acquire the relation between cause and effect of each node. By the interview with experts we can understand the influence of each factor on the SAR process. In order to construct the ES for SAR, we interviewed the workers in maritime search and rescue center, professors in SAR domain, and captain with experience.

Blind test: let expert to deal with different kind of SAR problem, and don't provide them with any SAR information. In the process of blind test we just give experts the SAR information which they asked. With this method we can describe dynamically the information which is need to inference, and understand the relationship of SAR information.

SAR cases: we can acquire SAR knowledge from SAR cases.

Other literatures: other researchers have worked for SAR knowledge; it is an important part of SAR knowledge acquisition, for example, the SAR knowledge in reference [5, 10, 12].

3.2 SAR KNOWLEDGE REPRESENTATION

After research the process of maritime search and rescue, we find that knowledge which was used to infer the SAR plan is fuzzy. For example, we provide the following knowledge for the rescue of man overboard.

If the weather is well, we should launch life boat to rescue the people in distress directly.

If the weather is bad, we should navigate the ship to downwind district of people in distress, and launch life boat just help him as a transit station beside the ship.

We generalized the rules in Table 2 from the rescue knowledge.

TABLE 2 Rules about man overboard rescue

Rule 65	IF THEN AND	The weather is well Launch safe boat Rescue the people in distress in distress
Rule 66	IF THEN AND	The weather is bad Navigate the ship to downwind district of the people in distress Launch life boat just help him as a transit station beside the ship

In Rule65 and Rule66, the "Well" set and "Bad" set of the weather are fuzzy set. For human experts, they can infer whether the weather is "Well" or "Bad". However, it is very difficult for expert system to judge the condition of weather from meteorological information. In SAR, and there are much knowledge like this. Therefore, we represent the SAR knowledge with fuzzy rules. Take the weather for an example. Assume the weather relates with wind, wave and rainfall, and we have the rule in Table 3.

TABLE 3 Rule about weather

Rule 50	IF OR OR THEN	Wind is wild. Wave is high. Rainfall is heavy. The weather is bad.
---------	--	---

Figure 4 shows the fuzzy sets of wind scale, and we discuss 3 fuzzy sets, which are Light (Wind is light), Medium (Wind is medium) and Wild (Wind is wild). Assume the scale of wind is 10, and then we can get formula (1), (2) and (3):

$$\mu(Wind = L) = 0, \tag{1}$$

$$\mu(Wind = M) = 0.2, \tag{2}$$

$$\mu(Wind = H) = 0.8, \tag{3}$$

where $\mu(Wind = X)$ refers to the membership degree of wind in fuzzy set X .

In the process of SAR, there are many fuzzy concepts, such as the level of grounding, the speed of sinking, and the level of damage. Not only the knowledge provided by experts is fuzzy, but also the SAR information we can acquire is fuzzy. That is to say, it is appropriate to represent SAR knowledge with fuzzy rules.

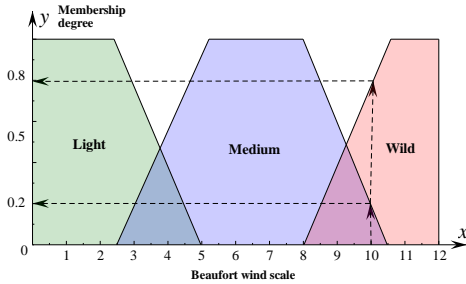


FIGURE 4 Fuzzy sets for the strength of wind

4 Inference engine of ES for SAR

The function of inference engine of ES for SAR is to generalize the SAR plan (results) from SAR knowledge (fuzzy rules) and SAR information (facts), the principle of inference engine of ES for SAR is shown in Figure 5. We can know that the inference engine is the core of ES, and relates with the SAR plan closely. It can deal with the SAR knowledge and SAR information. The former has been introduced, and then we would show the latter.

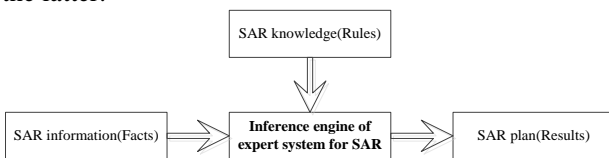


FIGURE 5 Inference engine of expert system for SAR

According to the source of SAR information, it can be divided into information of ship in distress, information SAR power (which means the ship or helicopter engaged in search and rescue), and information of meteorology and sea condition.

The work flow of inference engine is shown in Figure 6. The process of inference contains 4 steps, and they are respectively fuzzification of SAR information, evaluate the SAR information based on fuzzy rules, aggregate the output of rules, and inverse fuzzification of output.

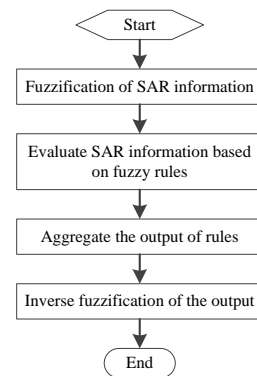


FIGURE 6 The workflow diagram of SAR inference engine

(1) Fuzzification of SAR information is to decide the membership degree of SAR information in appropriate fuzzy set. Fuzzy sets mentioned here corresponds to the fuzzy sets in SAR knowledge base, such as the fuzzy set about the strength of wind “Light”, “Medium”, and “Wild” sets. Generally speaking, the SAR information input into ES is a series of specific data. For example, the scale of wind is 9, height of wave is 3 meters, and flow velocity is 5kh. It is very difficult to match these data with fuzzy rules in ES, so fuzzification of SAR information is necessary. The key of fuzzification is to select the membership degree function. In order to improve the efficiency of ES the membership degree function is selected by experts.

(2) Evaluating the SAR information based on fuzzy rules is to match the SAR information with the left-side-part (LHS) of the fuzzy rules, and get the right-hand-side (RHS). If there is only one item in LHS, membership degree of RHS equals the LHS’s. For example, there is a rule in Table 4.

TABLE 4 A simple rule with one item in LHS

Rule 12	IF	Wind is light.
	THEN	Weather is Well.

And we have formula (4):

$$\mu(Wind = L) = 0.9 . \tag{4}$$

According to the principle, we have formula (5):

$$\mu(Weather = Well) = \mu(Wind = L) = 0.9 . \tag{5}$$

If the number of item in LHS is more than one, fuzzy operation (AND operation, OR operation) should be used when we evaluate the SAR information. For example, there is a rule in Table 5.

TABLE 5 A simple rule with more than one item in LHS

Rule 15	IF	Wind is Wild.
	OR	Wave is High.
	OR	Rain is Heavy.
	THEN	Weather is Bad.

And we have formulae (6-8):

$$\mu(\text{Wind} = \text{Wild}) = 0.6, \tag{6}$$

$$\mu(\text{Wave} = \text{High}) = 0.8, \tag{7}$$

$$\mu(\text{Rain} = \text{heavy}) = 0.4. \tag{8}$$

After fuzzy operation, we can get formula (9):

$$\mu(\text{Weather} = \text{Bad}) = \max\{0.6, 0.8, 0.4\} = 0.8 \tag{9}$$

(3) Aggregate the output of rules

There are so many rules about the same thing, so it is necessary to aggregate the output of rules. Fuzzy inference is used in ES for SAR, so the result of inference is series of membership degree of rescue operation in fuzzy set. For example, ES provides the following rescue method. Rescue method is to Launch safe boat(0.8), Rescue method is to use hook to clasp him(0.1), in which the value in parentheses refers to the membership degree. Generally, a SAR operation belongs to several fuzzy sets. In order to find the appropriate SAR plan, we should aggregate the output of rules.

(4) Inverse fuzzification of output

After aggregate the output of rules we get membership degree of rescue in several fuzzy sets. In order to decide which method should be taken, inverse fuzzification of the output is done. For common ES they would export a value, so centroid technique is used in inverse fuzzification of the output. Unlike them, ES for SAR need export a complete plan, that is to say we should know which fuzzy set the operation belongs to. So we use the maximum membership degree for inverse fuzzification.

5 SAR case

An example about rescue of man overboard is shown in this part to explain the inference engine of ES for SAR.

There are two rules shown in Table 6. In which x, y, and z (Weather, Man overboard, Rescue method) are variables in ES for SAR, A1 and A2 (Well, Bad) are fuzzy sets on field X (Weather), B1 and B2 (Near, Far) are fuzzy sets on field Y (Man overboard), C1 and C2 (Safe boat, Hook) are fuzzy sets on field Z (Rescue method). The membership function of A1, A2, B1, and B2 are shown in Figure 7.

TABLE 6 rules of Expert System for SAR

Rule 1		SAR Knowledge(Introduction)
IF	x is A1	Weather is <u>well</u>
AND	y is B2	Man overboard is <u>far</u> away from ship
THEN	z is C1	Rescue method is to Launch <u>safe_boat</u>
Rule 2		SAR Knowledge(Introduction)
IF	x is A2	Weather is <u>bad</u>
AND	y is B1	Man overboard is <u>near</u> from ship
THEN	z is C2	Rescue method is to use <u>hook</u> to clasp him

Assume $x=x1$ and $y=y1$, the process of inference would be shown in the following.

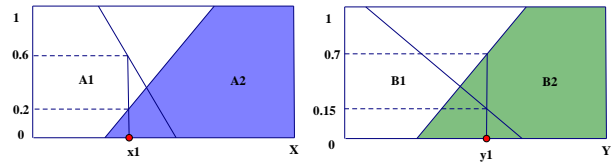


FIGURE 7 Fuzzification of SAR information

Step1: Fuzzification of SAR information. We can get formulae (10) and (11):

$$\begin{cases} \mu(x = A_1) = 0.6 \\ \mu(x = A_2) = 0.2 \end{cases}, \tag{10}$$

$$\begin{cases} \mu(x = B_1) = 0.15 \\ \mu(x = B_2) = 0.7 \end{cases}. \tag{11}$$

Step 2: Evaluate the fuzzy rules. After evaluate the fuzzy rules, we can get formula (12), and the result is shown in Figure 8.

$$\begin{cases} \mu(z = C_1) = 0.6 \\ \mu(z = C_2) = 0.15 \end{cases}. \tag{12}$$

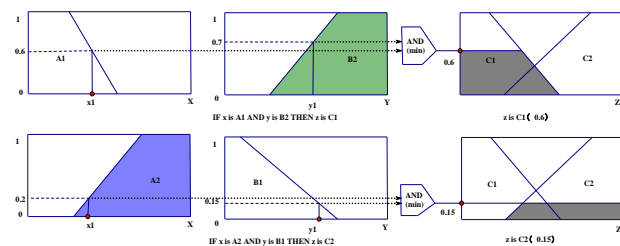


FIGURE 8 Evaluate fuzzy rules

Step 3: Aggregate the output of rules. The result is shown in Figure 9.

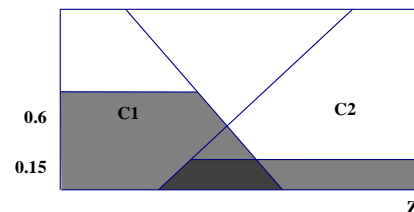


FIGURE 9 Aggregate the output of rules

Step 4: Inverse fuzzification of output of rules. The result is “z is C1”, that is to say “Rescue method is to Launch safe boat”.

The result of inference indicates that if weather is well, and the man overboard is far away from the ship, we should launch the life boat to rescue him.

6 Conclusion

ES for SAR can improve the efficiency of SAR, reduce the damage of life and property caused by search and

rescue delay. Compared with human SAR experts, ES can study new knowledge quickly and explain the whole process of inference. Additionally, ES can adapt complex work environment, and deal with large amounts of data. Due to the complex situation at sea SAR information is incomplete. Fuzzy rules are used to represent SAR knowledge. On the basis of fuzzy rules this paper has designed the structure of ES for SAR, and researched the SAR knowledge acquisition, representation and inference engine.

Acknowledgements

This work was financially supported by the National Nature Science Foundation of China Project (No. 61073134, No. 51179020), National Natural Science Foundation Project of China (Grant NO. 51309044), Scientific Research Project of Liaoning Education Department (Grant NO.L2013208), and the Applied Fundamental Research Project of Ministry of Transport of China (NO. 2013329225290).

Reference

[1] Xinyuan T, Hongzhang D, Zuoxiong Z 2000 Development Policy of SAR Communications in China *Navigation of China* (1) 51-5 (in Chinese)

[2] Haobing N, Shengzheng W, Zhiwu H, Chaojian S 2011 Dynamic optimizing ship routing algorithm for applications in maritime search and rescue *Journal of Shanghai Maritime University* (4) (in Chinese)

[3] Fangbing X 2011 *Research on the Key Technologies of Maritime Search and Rescue Decision Support System* Dalian: Dalian Maritime University (in Chinese)

[4] Shengwei X 2012 *Research on Global Optimization Model and Simulation of Joint Aeronautical and Maritime Search*. Dalian: Dalian Maritime University (in Chinese)

[5] Zhiwei G 2011 *Research on Water Emergency Rescue Decision Support System Based on Case Based Reasoning*. Dalian: Dalian Maritime University (in Chinese)

[6] Cottam H, Shadbolt N 1998 *International Journal of Human Computer Studies* 48(4) 449-73

[7] Cheng S-C, Tsai W-Z, Chen Y-Z 2008 Rescue knowledge M-learning system by 3G mobile phones *5th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education* 148-52



[8] Cheng S-C, Tsai W-Z, Chen Y-Z 2008 Information technology in rescue knowledge learning system *IEEE International Symposium on IT in Medicine and Education 2008* 469-74

[9] Van Dyne M, Tsatsoulis C 2012 An inferential system for determination of candidate crash sites for search and rescue operations *IEEE Aerospace Conference 2012* 1-9

[10] Shengyou L 2011 *Research on the Knowledge Base Construction and Reasoning Mechanism of Salvage Expert System* Dalian: Dalian Maritime University (in Chinese)

[11] Yunusoglu M G, Selim H 2013 A fuzzy rule based expert system for stock evaluation and portfolio construction *An application to Istanbul Stock Exchange* (1) 908-20

[12] Run P 2012 *Research on Knowledge Management in the Field of Maritime Search and Rescue Based on Agent*. Dalian: Dalian Maritime University (in Chinese)

Authors	
	<p>Zhang Yingjun</p> <p>Current position, grades: Director of Research Institute of Traffic Information Engineering in Dalian Maritime University, the senior member of China Institute of Navigation, Professor, Ph.D. in Technical Sciences.</p> <p>University studies: Admiral Makarov State Maritime Academy (1990-1995), Ph.D on Technical Sciences.</p> <p>Research interests: large information system development method; decision theory and method; optimization model and algorithm</p>
	<p>Yang Xuefeng</p> <p>Current position, grades: Ph.D. student of Navigation College of Dalian Maritime University</p> <p>Research interests: intelligent transportation systems</p>

Research on dynamic risk identification model of shield tunnelling based on REASON model

Jun Luo^{1*}, Sijing Cai¹, Yanhui Wang²

¹Civil and Environmental Engineering School, University of Science and Technology Beijing, No.30,Xueyuan Road, Haidian District, Beijing, China

²School of Traffic and Transportation, Beijing Jiaotong University, No.3, Shangyuan Village, Haidian District, Beijing, China

Received 6 May 2014, www.tsi.lv

Abstract

Aiming at the dynamic risk identification problem in shield tunnelling, and with the lack of research on dynamic risk identification theory and human factors in shield tunnelling, an analysis model of shield tunnelling based on REASON model has been proposed to establish in this paper. Relying on the fault tree theory and the model that established, the accident rule base has been built. After forming the REASON model into a network, the dynamic risk identification model for shield tunnelling has been built to provide theoretical guidance for dynamic risk management during the construction.

Keywords: Shield tunnelling, REASON model, dynamic risk identification model, risk management

1 Introduction

With the accelerating of urbanization process in china, urban population is expanding rapidly and land resources are becoming scarce. In addition, the conflict between the growing traffic demand and the increasing congestion of urban ground traffic has become particularly sharp. Aiming at this, development and utilization of city underground space has become an inevitable trend and an important means. By the end of 2013, metro has been operated in more than 19 cities in Chinese, and the total mileage has been up to 2366 KM [1]. The metro construction projects are in a stage of rapid development in scale and quantity, as shown in figure 1.

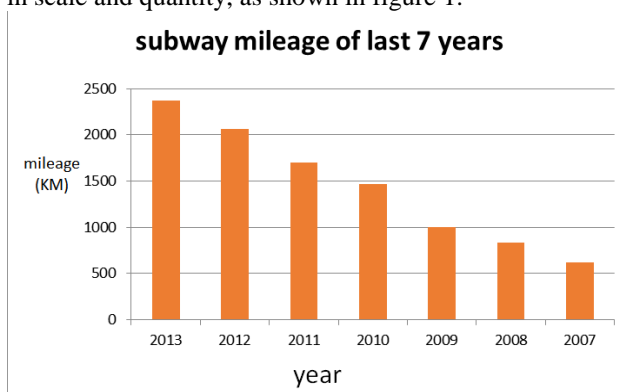


FIGURE 1 Statistics of metro mileage in last 7 years in China

According to the statistics, 88% of the accidents in the metro construction process are caused by unsafe behaviour of human, 10% are caused by the effect of both unsafe behaviour of human and material insecurity status, only about 2% are caused by non-human factors. Also, most of the accidents that caused by unsafe behaviour of

human are due to the fault of construction management [2-3].

Although, some progress has been get aiming at the problems of safety management, problems are still existed as followed:

(1) Research on risk assessment and management in shield tunnelling is still concentrated in the aspects of overall risk evaluation and reliability calculation, which is lacking for dynamic risk assessment and has not formed a set of management system. In addition, the achievements are out of practice, which cannot be used in practical projects.

(2) As the lack of theory research on human factors, importance is also lacking in practical projects, and deeply, the potential organizational factors that affect human behaviour have been ignored.

Therefore, in view of the questions above, research on risk analysis and risk identification of shield tunnelling has become increasingly urgent. Based on the risk identification of shield tunnelling, deeply influence of organizational factors on human has been analysed in this paper. By researching on the dynamic risk management theory from systematic perspective, the objective that improving management efficiency, avoiding engineering risk and improving economic benefit should be realized.

2 REASON model

REASON model is a conceptual model that proposed in a book called "Human error" by James Reason, who is a professor in University of Manchester. The core innovation point of this model lies in that from the perspective of system theory, unsafe behaviour of human and its potential organizational effect factors have been

* Corresponding author e-mail: author@domain.com

analysed. In the view of direct and indirect impact among managers, stakeholders and corporate culture, the perspective of accident analysis has been all-round expanded, and also all the relevant factors have been catenated by a logical accident chain reaction [4-5].

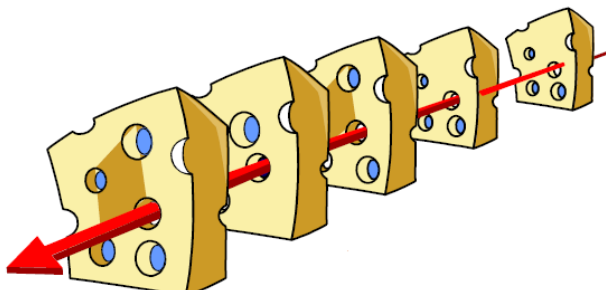


FIGURE 2 Original Reason Model

The original REASON model that shown in fig.2 includes the following aspects of meaning:

(1) As each piece of cheese represents an event, every hole of the cheese means a failure point. When a straight light can pass all cheese through the holes, accidents will occur.

(2) As long as moving a piece of cheese, that makes the light cannot penetrate, the accident can avoid.

(3) The model emphasizes the overall crash prevention ability of the organization. The core of the model is how to minimize the cost of management for maximum benefits, which means just prevent an accident from a piece of cheese, rather than for all defects of every piece of cheese.

After the proposing of REASON model, the model has been appropriate revised by researchers from various fields, such as man-machine engineering, medicine, nuclear industry, and aviation, marine, in order to reduce accidents. Up to now, many revising models have been proposed which control the occurring of accident effectively [6-7].

The introduction of REASON model into shield tunnelling safety management can help to establish the risk analysis model and find the fault chain for effectively prevention of accidents. But, because of the limitations of traditional REASOM model and the improving models, the application of them in shield tunnelling risk management still exist following problems and should be improved.

(1) Limited by researcher's field, the application field of existing improving reason model is also limited and may not applicable for other fields. Aiming at the complexity, maintainability and multi management system of shield tunnelling, the model must be improved particularly.

(2) The internal logic of REASON model shows that, accidents may happen when a straight light can pass all cheese through the holes. But in shield tunnelling projects, vulnerability can even cause an accident which may bring huge loss.

3 REASON model of shield tunnelling

3.1 HIERARCHY OF THE MODEL

The improved shield construction REASON model is consist of six layers, like planning and decision layer, safety supervision layer, organization management layer, dangerous premise layer, unsafe behaviour layer and the recovery layer. Among of the six layers, the human error of planning and decision layer, the safety supervision layer, and the organization management layer will not directly lead to risk events, and these layers belong to invisible factors level. The dangerous premise layer includes the self-situation of construction operations staff, as well as environmental factors, the failure of which may directly lead to risk events. Therefore, the dangerous premise layer is part of semi dominant factors. Unsafe behaviour layer and the recovery layer are belonging to the dominant factors level. Each level is described as follows.

(1) Planning and decision layer

Risk has already existing when a shield tunnelling project is in the stage of planning and design. For example, the place and method that chosen to construct the tunnel may potential influence on tunnelling construction. Generally speaking, the human errors of this layer contain programming errors, design failure, major decision errors, the fail subject of which are management departments and design departments.

(2) Organization and management layer

The most harmful errors are the invisible errors that latent and far from accidents, which usually exist in the organization and management of construction. When the behaviour that may bring about accidents is not existing, failure of the management layer will not make a threat of construction safe, which cannot be found easily. The organization and management layer mainly includes the company culture, organization structure, drawings, construction scheme, training management, operation management and resource management.

(3) Safety supervision layer

Safety supervision layer mainly emphasizes on safety supervision in the process of construction, which includes controlling and managing persons on site in the construction unit, monitoring unit's daily supervision, construct unit and all kinds of inspection of higher competent department. The human errors of this layer are mainly divided into four aspects: improper supervision plan, inadequate supervision, uncorrected known problems and supervision violations.

(4) Dangerous premise layer

Dangerous premise layer is made up of self-situation of construction operations staff, construction environment and equipment safety situation. Self-situation of construction operations staff includes psychological, physiological, skills, knowledge of line workers.

(5) Unsafe behaviour layer

Unsafe acts mainly refer to the direct behaviour that line workers done to cause accidents. It is consist of fault and illegal.

Fault is an unintentional behaviour, which means the inappropriate behaviour of line workers that deviated from willingness or construction scheme. Fault mainly includes four kinds of error: perceptual error, memory error, decision-making error and skill error.

Illegal means the violation of rules and operating procedures. It mainly contains intentional violation and unintentional violation. The unintentional violation is generally happens in the case of unfamiliar with the regulations, but the intentional violation happens in the case of that line workers know clearly that their operation is inconsistent with the rules and procedures.

(6) The defence and recovery layer

This layer is mainly means the measures after the risk factors are found unsafe.

As mentioned above, the improving shield construction REASON model as follows.

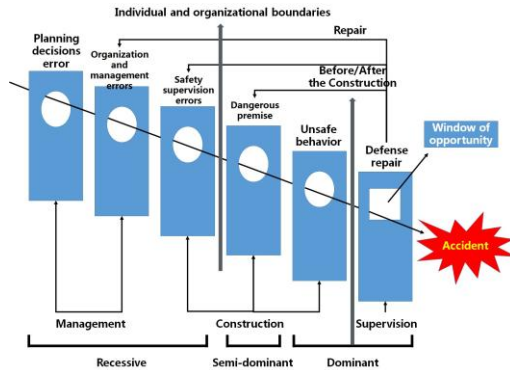


FIGURE 3 The improving REASON model of shield tunnelling

3.2 HIERARCHY OF THE MODEL

3.2.1 Common accident statistics of shield tunnelling

As Shield tunnelling construction system is composed of geology, shield machine and human, the shield tunnelling accidents are caused by geological, machine and man-made reasons. From the statistics, the major accident categories are shown as following table.

TABLE 1 Common accidents of shield tunnelling

No.	Name of the accident	No.	Name of the accident
1	TBM broken parts	11	Floating of tunnel
2	Damage of shield cutter	12	Twist of TBM
3	Mud cake	13	Gush
4	Fracture of jack	14	Overrun of boring deviation
5	Fracture of lifting head of segment installation	15	Starting plunge of TBM
6	Clogged pipeline	16	Segment broken
7	Accidents of circuit and pipeline	17	Jam of TBM
8	Subsidence and uplift of ground	18	Fires
9	Tilt and damage of house	19	Personal injury accident
10	Gushing of water and sand		

3.2.2 Risk identification of shield tunnelling

The way used to identify the risk of shield tunnelling construction in this paper is a method that combines work breakdown structure method and risk breakdown structure method.

(1) Analysis of shield construction process

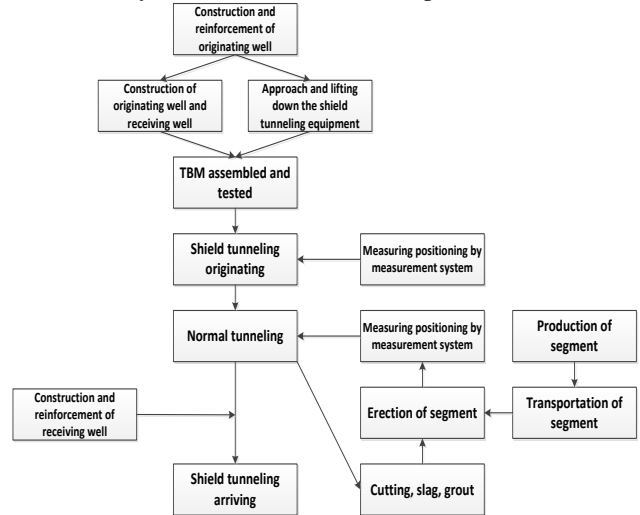


FIGURE 4 Major processing flow of shield tunnelling

(2) Overall risk analysis

According to the scope of this paper and the shield main flow, the overall risk has been analysed as shown in table2.

TABLE 2 Main analyses of risk identification in shield tunnelling

Environment	
Technical risks	Construction Preparation, Shaft Construction, Measurement, Shield Tunnelling, Earthwork Cutting, Slag, Segment Erection, Simultaneous Grouting, Secondary Grouting, Waterproofing And Drainage, Originating, Arriving, Lifting Across the Rivers, Formation Empty, Quicksand, Gas Layer
Geological risks	Earthquake, Cold, Typhoon, Rainstorm, Flood
Natural risks	Nearby Buildings, Obstructions Within The Formation, Pipeline, Nearby Existing Traffic
Surrounding	Construction Organization, Construction Management and Organization risks
Environmental risks	Personnel Management, Device Management, Fire Management

(3) Specific risk analysis

Based on general analysis, the 185 risk factors of shield construction are analysed by the way of list.

3.2.3 Elements extraction

Aiming at the demand of risk management in shield tunnelling, combined with the analysis of risk factors, the elements of the model have been extracted as shown in figure 5.

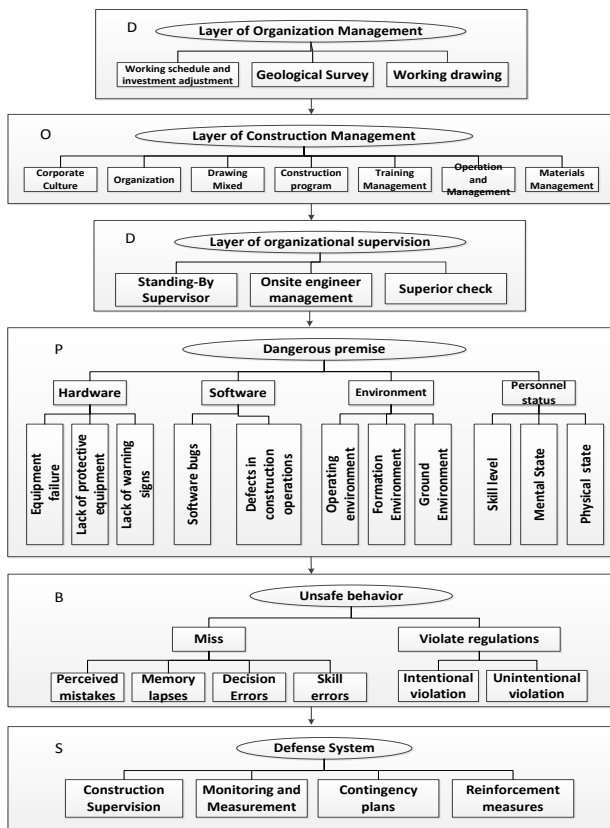


FIGURE 5 Elements system of Reason Model in shield tunnelling

3.3 ACCIDENT CAUSATION RULE BASE BASED ON FAULT TREE

There is a great number of factors that can affect shield tunnelling safety, the relative importance of which may changes at different time. In order to discern the pattern of tunnelling accidents from an overall perspective, fault trees of shield tunnelling have been built, which can not only arrange the relationship among the factors, but also help to set up an accident causation rule base as shown in figure 6.

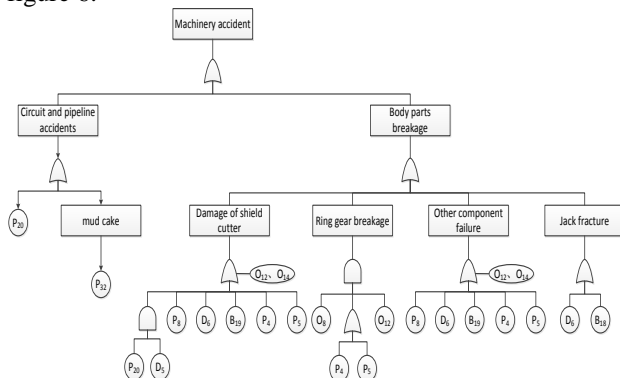


FIGURE 6 Event tree of mechanical failure

4 Dynamic risk identification of shield tunnelling

4.1 NETWORK DESCRIPTION OF REASON MODEL

4.1.1 Basic definitions and assume

Hypothesis 1: The network consists of the risk factors (node) and their relationship (side) of REASON model.

Hypothesis 2: System topology is fixed.

Definition 1 Node v: Suppose a point $v \in V(G)$, $v = \{0U1|P, C, H\}$, V contains only two kinds of status: 0 or 1, and v contains three properties: p (failure rate), c (loss degree), and H (Human factor). Then V is called a node. Set of nodes is $V(G)$. Nodes have two types, called conditional node and unconditional node, as shown in figure 7.

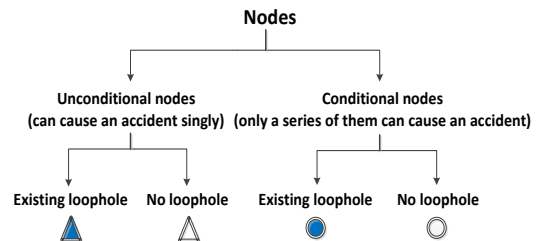


FIGURE 7 Classification of the nodes

Definition 2 Failure rate p: It is the description of the existence probability of loopholes in the nodes of the network.

Definition 3 loss degree c: Description of the potential loss that the nodes may bring.

Definition 4 Human factor h: Description of the influence degree that human state may affect the node status.

Definition 5 Interaction relationship e: Assuming that any direction line segment $e=(u, v)$, in which $u, v \in V(G), u \neq v$; e is named interaction relations, means elements u, v can simultaneously have an effect on some others. The set of mutual relations is $L(G)$.

Definition 6 cause relationship f: Sets the direction line $f=(u \rightarrow v), u, v \in V(G), u \neq v$, calling f cause relationship, means U occurs will lead to v . The set of cause relationship is $F(G)$.

Definition 7 Domain G: Assume $G_i = \{V_i(G), E_i(G), F_i(G)\}$; G_i is called a domain.

Definition 8 Set domain G: Assume $G = [G_1, G_2, \dots, G_n]$, G is called a set domain.

Definition 8 Set domain G: Assume $G = [G_1, G_2, \dots, G_n]$, G is called a set domain.

4.1.2 REASON model network

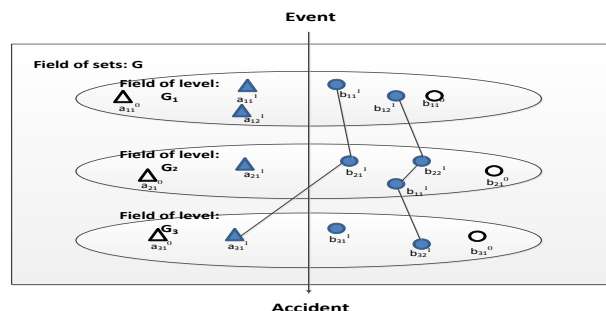


FIGURE 8 Schematic of network model of risk assessment in shield tunnelling based on Reason Model

(1) Risk factors constitute the basic unit of the model is divided into elements and no loopholes exist two types of vulnerability factors. With solid point v_{ij} and hollow point v'_{ij} said, $(i=1,2,\dots,n, j=1,2,\dots,m)$, $v_{ij} \cup v'_{ij} = V(G)$;

(2) Elements can be composed according to certain rules of multiple domains represent different management levels. G_1, G_2, \dots, G_n represent an elliptical area.

(3) Among the different elements of the same domain and different domains, there are two kinds of relationships. Namely the interaction relations and cause relations, expressed in e_{ij} , f_{ij} respectively.

(4) Probability p elements in attribute changes between 0 and 1, 0 properties for solid, 1 attribute is hollow.

(5) G_1-G_n together constitute the domain security management feature set $G, V(G) \cup E(G) \cup F(G) = G$;

(6) When there is a T occurs, the events of the line across the G line may be a straight line, also possible tree bifurcate structure;

(7) Domain G_1 to G_n all within the G movement according to certain rule, T intersection with the event circumstances exist that may occur during the movement. The law of the domain and the event T intersection of influenced by the interaction and cause relationship.

(8) For the occurrence of any event, then the intersection of the event elements and fields are holes will lead to accidents. Assume $v_i^* = T \cap G_i$. If $v^* = \{v_1^*, v_2^* \dots v_n^*\} \neq \emptyset$, and $\forall v_i^* \in v^*, \exists v'_{ij} \in G_i, v'_{ij} = v_i^*$, and accidents $A \neq \emptyset$.

4.2 DYNAMIC RISK IDENTIFICATION MODEL

The realization logical of the model is that when an event occurs, elements vulnerabilities are identified according to the order of the domain layer. Identification procedure is as follows:

Step1: Recognizing the loopholes of unconditional nodes in the first layer one by one, if existing, making a record and go on, if not existing or all nodes have been checked, go to next step.

Step2: Recognizing the loopholes of conditional nodes in the first layer, if existing, based on the relationship between the interaction relations and cause to track identification, identify vulnerabilities propagation

References

[1] National bureau of statistics of the people's republic of China 2007-2013 Chinese statistical yearbook
 [2] Xie D S, Wang M Y, Lu H 2010 Accidents statistics and analysis of metro engineering *Symposium of the second session conference of national construction safety and protection* 205-9
 [3] Deng X P, Li Q M, Zhou Z P 2010 Statistic and analysis on accident regularity of shield tunnelling *Statistics and decision* 87-9

path, skip to Step3; If not present, then the layer marked as safe, vulnerability factors attributes also change accordingly, and then jump to the Step4.

Step3: Contrasting identifies vulnerabilities and incidents caused by the propagation path of the rule base, if matched, and then there might be an accident, if not match, then safety.

Step4: Cycling step1, step2, and then the remaining layers sequentially domain identification, if the layers are marked as safe, then the overall safety.

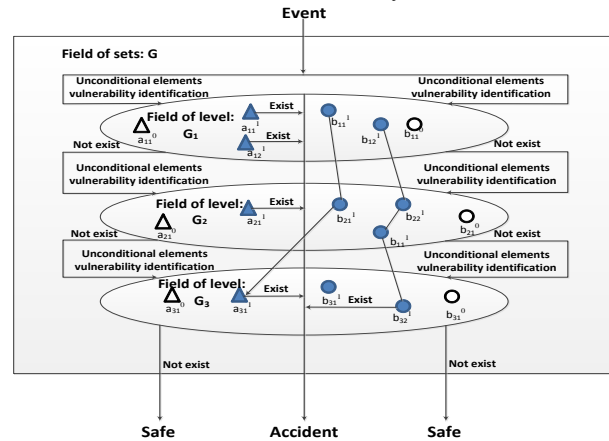


FIGURE 9 The logic of dynamic risk identification

5 Conclusions

Shield construction process is a complex system, which is more than one department, multidisciplinary, and multiple factors. And also risks exist in the dynamic changes which need to be dynamic and static analysis based on traditional risk assessment. Based on the traditional REASON model analysis, combined with the characteristics and risk factors shield construction accident itself, to establish a shield construction REASON mode, and model layers of elements were extracted; On the basis of the REASON network model described in the proposed shield construction dynamic risk identification model to provide theoretical guidance for identifying dynamic risk shield of the construction process.

[4] Reason J 2010 Human error *Cambridge University Press*
 [5] Liang G, Zhang Y J, Li H, et al 2008 Quasi-reason model for human errors analysis in substation operation *Relay* 23-6
 [6] Wu W Z, Yang L 2011 Maritime investigation model based on Reason Model *World shipping* 42-5
 [7] Chen N T, Tan X, Li R 2012 Application of REASON Model to Investigation of the Aviation Maintenance Accident *Computer and Communication* 96-8

Authors	
	<p>Jun Luo, born on July 14, 1986, Nanchang, Jiangxi province, China</p> <p>University studies: M.Sc. in Beijing Jiaotong University (2010). Is studying for PhD in the University of Science and Technology Beijing. Scientific interests: include different aspects of mining safety management and traffic construction safety management.</p>
	<p>Sijing Cai, born in October, 1952, Nanchang, Jiangxi province, China</p> <p>Current positions, grade: professor and tutor of doctorate candidate of University of Science and Technology Beijing. University studies: M.Sc. in the University of New South Wales in Australia and PhD in the Kyoto University of Japan. Scientific interests: safety technology and Engineering.</p>
	<p>Yanhui Wang, born on April 20, 1974, Baoding, Hebei Province, China</p> <p>Current positions, grade: Associate professor of Beijing Jiaotong University. University studies: M.Sc. and PhD University of Science and Technology Beijing. Scientific interests: safety technology and Engineering of traffic.</p>

Tag-based process recommendation for social business process modelling

Yanming Ye^{1, 2*}, Yueshen Xu², Zhilin Feng³

¹College of Information Technology, Hangzhou Dianzi University, 310018 Hangzhou, China

²College of Computer Science and Technology, Zhejiang University, 310027 Hangzhou, China

³College of Zhijiang, Zhejiang University of Technology, 310024 Hangzhou, China

Received 12 May 2014, www.tsi.lv

Abstract

Social BPM (Business Process Management) has become a new research hotspot in business process management field because of its capability of handling the flexibility and dynamics of process in social circumstance by means of integration of social software and BPM. The key technique is process modelling, and note worthily the process modelling is more complex in social BPM than in traditional BPM. This paper presents the definition of social business process model to identify the difference with the traditional business process model and gives a prototype of social BPM system. The modeller in the prototype system may be a common user without professional knowledge, so a tag-based process recommendation method is proposed to facilitate modelling. The experiment result shows that the method is valid and effective in computer-aided intelligent process modelling.

Keywords: process recommendation, process modelling, social business process model

1 Introduction

In traditional Business Process Management System (BPMS), process modelling is mainly conducted in a certain sequence via some modelling languages like BPMN (Business Process Modelling Notation). Further, those intermediate results are arranged and combined together by a process engine using a flow-oriented paradigm [1-3]. With popularity of Web 2.0 technologies, more and more business applications are deployed in form of Web applications, which lead to the socialization of business processes to better reach users and reduce the total cost. As a result, these business processes become more flexible and collaborative. Meanwhile, their activities become long running loose-coupled [4-7], which brings a challenge to traditional BPM systems that can only support the well-structured processes. Therefore, recently, people are making much endeavour to modify the traditional BPM to be adapted in new circumstances. Especially, social BPM techniques are getting more and more attentions for its combination of BPM and social software, while the latter can support social interaction and production, and raise the interaction level and scope, which are facilitated by computer networks [8-10].

In social BPM, social applications are integrated into business process management. Wherein, the participation of external stakeholders is strictly controlled in process design and implementation. To make social business more flexible and dynamic, process modelling may become more important and complex, so intelligent

modelling methods are required by most of social BPM systems.

This paper focuses on the computer-aided process modelling of social business process. For clarity, JTangSBPMS that is the basic social BPM system of this paper is introduced firstly in Section 2. Then, in Section 3, we give some definitions with related basic instructions. Next, the process recommendation method is discussed in detail in Section 4 and experiments are demonstrated in Section 5. Finally, we show the conclusion and future work.

2 JTangSBPMS

Since social applications are prevailing in the whole world, the combination of microblog with existing BPMS is become necessary. Therefore, we build JTangSBPMS, a new social BPMS, whose prototype is shown in Figure 1.

As the same with traditional BPM, the lifecycle of social BPM can also be decomposed into the four following phases: design, configuration, enactment and diagnosis. In JTangSBPMS, process modellers and task performers are all microblog users. Besides, to complete the task assignment, a message containing the short URL of web application is sent to the target user, since the application owns the task widget. Then, the URL should be clicked to open the Web application, which will be deployed automatically. Users can send messages in the microblog about their feelings and experiences on the task or the business. The user collaboration can be

* *Corresponding author* e-mail: yeym@hdu.edu.cn

realized by the information exchanging through message sending.

In design phase, with the increasing of flexibility and collaboration of business process, process modelling also becomes more important. Process modelling is basic of the whole lifecycle. For the complexity, experts or professionals usually conduct the process modelling in traditional BPM systems. However, in social BPM systems or collaborative environment, the users or workers are peer to peer and the possible applied business areas are very wide, so the process modelling must be oriented to common users. Because the correctness and completeness of a process model is important for business process, it is necessary to provide intelligent modelling method to eliminate much of the hardship. This paper presents a tag-based process recommendation method to realize the intelligent modelling in JTangSBPMS.

The rest phases of JTangSBPMS are all based on the social business process model that is modelled in the design phase. For the configuration phase, the key steps are to configure Web applications carrying the real tasks according to activities of the process model. When an activity is on execution, the short URL of the Web application that carries the task of the activity will be sent to the performers in form of microblog message. In the enactment phase, any dynamic change on the business process will launch the process verification and recommendation for the robustness of the process. In each time when a process is finished, a social business process execution model will be produced, which can record the real running data. As for in the diagnosis phase, there are various social network analysis tools or methods that can be used to analyse the user's behaviour and process structure as well as other social features of the process. Finally, a social business process can be improved based on these analyses.

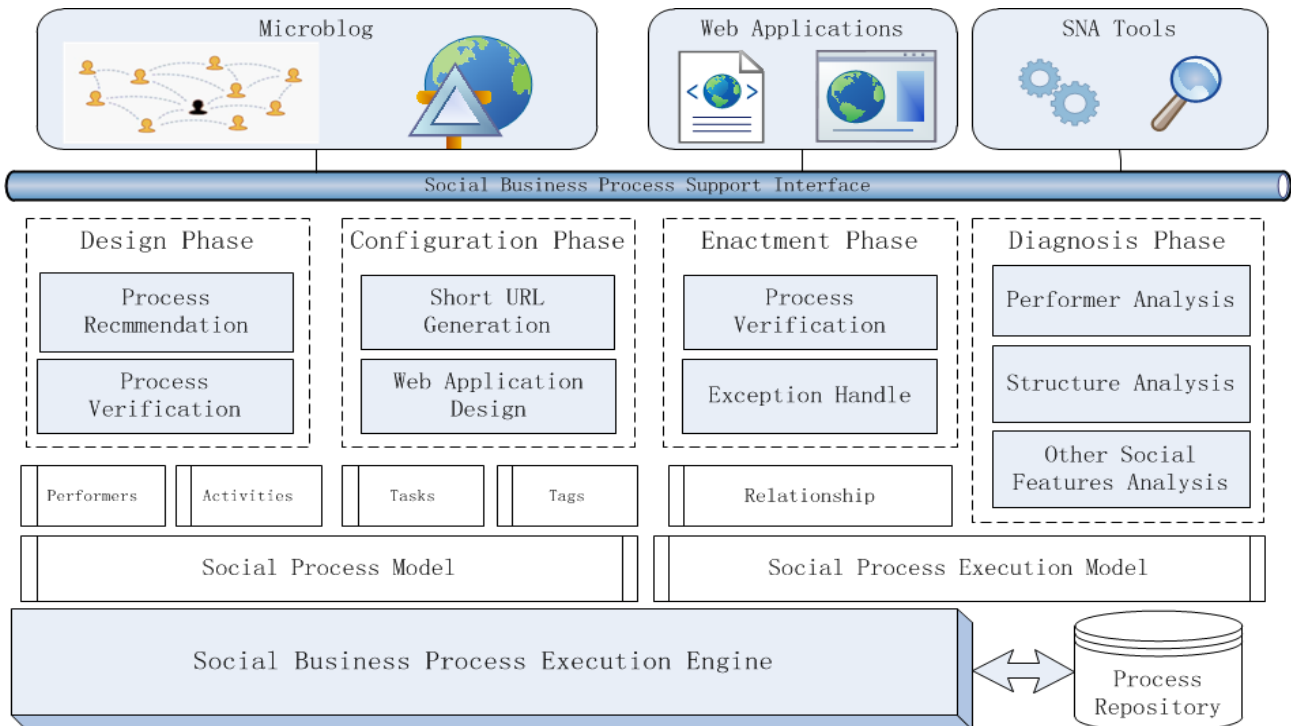


FIGURE 1 Prototype of JTangSBPMS

3 Preliminaries

From the above discussion, the traditional BPM is flow-oriented, and business process tasks are scheduled in a specific sequence. Such an activity has a certain position in a process model, and the process is well-structured so that it can be executed in the pre-defined orders of activities in enactment phase. Meanwhile, the social BPM must deal with flexible and collaborative features of processes. For better comprehension, we will begin with the definitions of traditional business process model and social business process model.

Definition 1. Traditional Business Process Model:

Let (P, T, F) be a WF-net (as defined in [11]), and a traditional process model is a 5-tuple (P, T, F, U, Ω) where,

- (i) U is a performers set.
- (ii) $\Omega: U \rightarrow P$ is the assignment function of performer (U) to a place (P) that includes certain task.
- (iii) $P, T,$ and F represent places, transformation, and flow structures respectively.

Definition 2. Social Business Process Model: A social business process model is a 5-tuple (C, U, ξ, R, Ψ) where,

- (i) C is an activity set in which each activity contains a specific task.
- (ii) E is a user set as the subset of the users in Definition 3.
- (iii) $\xi: C \rightarrow \{subset(U)\}$ is a function which maps activity to possible performers who are subset of the whole users.
- (iv) R is a finite set containing five different kinds of relationships and $R = \{\text{before, after, and, or, undefined}\}$.
- (v) $\Psi: C \times C \rightarrow R$ is the relationship function between users.

From the above definitions, we can see that most of traditional process models are flow-oriented and the activity performer is usually appointed a beforehand. But in the social process model, the order of activities may not be strict and we can determine the performers until running time.

Next, a formal definition of social network will be given as follows.

Definition 3 (Social Network): A social network SN is a graph denoted by a 5-tuple $SN = (N, T, M, R, K)$ where,

- (i) N is the finite set of nodes and each node stands for a user.
- (ii) T is a set of social tags.
- (iii) $M: N \times T \rightarrow \{true, false\}$ is the mapping function between users and tags
- (iv) $R = \{R_1, \dots, R_m\}$ is a finite set of relationships.
- (v) $K: N \times N \rightarrow R$ is the relationship function.

In a social network, a user's preference for a resource or a task can be inferred by tags, which can reflect the user interest and capability. Usually, the increasing of the number of tags means that a user is greatly interested in a certain resource or a user has many capabilities on a certain task.

Although the social business process model may be ill-defined, for example, the order of activities and the performers of activities are often unclear, any model execution will be conducted certainly on the unclear features in running time. Therefore, we give the definition of the process execution model as follows.

Definition 4. Social Business Process Execution Model: Let Y be a set of activity types and B be a finite set of tags for activities. A social business process execution model is a connected graph denoted by a 5-tuple $P = (S, E, D, \alpha, \beta)$ where,

- (i) S is a finite set containing activities.
- (ii) E is the finite set of users
- (iii) $D \subseteq S \times S$ is the finite set of edges.
- (iv) $\alpha: E \rightarrow S$ is the users assignment function for activities.
- (v) $\beta: S \rightarrow B$ is the activity tagging function.

It is obvious that the social business process execution model is much like the traditional business process model in which each component is clear, including performer and trigger condition of every activity and the

relationship of the activities. Therefore, we can analyse the execution models to improve the social process model or create a new social process model. In a social network, the users are naturally clustered according to their interests and work experiences and the clusters are reflected in the tags that the users hold. In our system, each process model and its each activity will be labelled with different social tags depending on features of the business and the task, which are named *process tag* and *activity tag* respectively. If a user successfully models a social business process and publishes the model, the user node will automatically send a message containing the *process tag* and *process name*. In the application, if a user accepts and completes a task, the user node will automatically send a message containing the activity name and process name with their related tags.

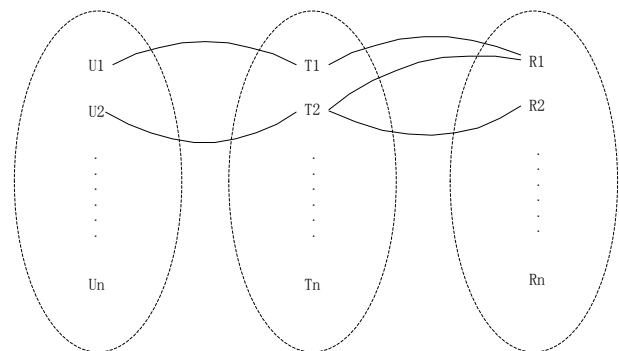


FIGURE 2 Formal representation of tagging by the triples defined in the users, tags, and resources spaces

In our system, the tags are divided into two types that are *process tag* and *activity tag*. There are three main components for any tagging system: a set of users, a set of tags created by the users and resources (URL, entities etc.). A social tagging system can be represented by a triple of users, tags and resources: (U, T, R) , and the conceptual model of a tagging system is shown in Figure 2. [12-14]. Resources are restricted to be processes and activities depending on the specific purpose of our system in this paper. A tag is actually a word or a phrase that expresses certain meaning. However, people could use a wide variety of words to describe the same objects. Especially in the case of social tagging systems, users may apply “different terms as tags to describe the same resource” by using synonyms, homonyms and polysemy, leading to multiple and diverse descriptions for the same resource. The increasing number of vocabularies may imply that the connections between tags and resources will become less and less distinct, making information retrieval more difficult. To resolve the semantic problem, our system adopts an eclectic solution that provides a recommendation tag for each process model or activity based on the tags dictionary. Users can accept the recommendation default tag without any change or change it when they edit the related profiles or messages. If the users change a tag that is recommended, the tag will be added into the tags dictionary in proper catalogue unless it exists in the dictionary. The tags dictionary can

be trimmed from the WordNet Domains by means of some improvable methods [15-16]. While for simplicity, in our system the tags dictionary is in form of (tag, process-name/activity-name), of which the process tags are clustered by the probabilistic latent semantic analysis [17] on the set of the processes in the process repository. The activity tags are some fungible domain labels that are obtained from the WordNet Domains.

To model a complex social business process, it can be simplified as to confirm the next activity and its performer as clearly as possible, and then to specify the relationship between the activity and the existing activities of the modelling process fragment step by step. In general, two process models that belong to the same business area and resolve the similar business problem have the similar structure. A user may be more professional or skilful to perform a task if he or she has performed the similar task more times or more than other tasks. In a collaborative environment, a user mostly tends to work with the other certain user or users that he or she is familiar with. Therefore, we can recommend the next activities and their performers when a user is modelling a business process based on the historical process models and user behaviours. As what has been analysed, our recommendation method in fact is to find out process models whose structures are similar with each other and users whose behaviours are similar with each. Then, we will recommend the next activity that exists in most similar process models after the sequential activities that are contained in modelling process fragment, with the performers as users that have similar behaviour.

Definition 5 (Process Code): The process code of a process model (or process execution model) can be represented as the breadth first traversal order of tags on the process graph, which is a linear order, where the tags of the same hierarchies are by lexicographic order.

For example, the process code of the process execution model *p*, as show in Figure 3, is “ABCDEF” where the “A”, “B”, “C”, “D”, “E” and “F” are tags of activities of the process, not “ABDCEF” or others.

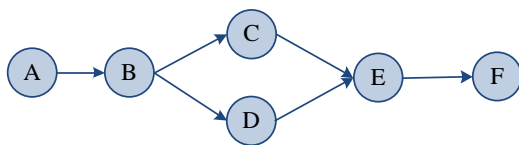


FIGURE 3 A sample of a social business process execution model -p. The “A”, “B”, “C”, “D”, “E” and “F” can be seen as the tags of the activities that the users have labelled

It is necessary to notice that the process code may be an array of chars or a string that can convert one to the other easily.

Definition 6 (Structure Similarity): the structure similarity between process execution model *ep₁* and *ep₂* with the process codes as *cep₁* and *cep₂* respectively can be calculated by the following formula,

$$Sim_structure(ep_1, ep_2) = \frac{|cep_1 \cap cep_2| + \sum_{\substack{ti \in (cep_1 - cep_2) \\ tj \in (cep_2 - cep_1)}} synonym(ti, tj)}{\max(|cep_1|, |cep_2|)}$$

and $|cep_2|$ is the length of *cep₁* and *cep₂* respectively, and the *synonym(x, y)* is the synonym function that returns 1 if the given tags *x* and *y* are synonyms and 0 otherwise.

Performers are important elements in a social business process, whereas many modellers have some doubts on how to assign more appropriate users to the tasks. The performer recommendation can facilitate the procedure, and the tag weight must be calculated in first.

Definition 7 (User Behaviour Similarity): The user behaviour similarity of a user on a tag can be calculated as the following formula,

$$Sim_behavior(u, t) = tf(u, t) / \sum_{i=1}^n tf(u, t_i) \quad , \quad \text{where}$$

tf(u, t) is the tag frequency that represents the times that a tag *t* has been tagged by a user *u*, and *n* is the total number of tags that user *u* has bookmarked.

4 Implementation

Tag-based process recommendation method turns to help process modellers fit processes to achieve a modelling intention with regard to the other modellers’ modelling behaviour and achievement.

The most important step of the method is to implement the algorithms of process structure similarity and user behaviour similarity. Next, we give the pseudo code of the structure similarity algorithms (shown in Figure 4).

Input: Process code: *cep₁*, *cep₂*

Output: Process structure similarity: *structuralSim*

Declaration:

String[] intersect(String[] arr1, String[] arr2)

String[] difference(String[] arr1, String[] arr2)

int synonym(String arr1, String arr2)

```

1: Initialize: structuralSim=0, int_syn=0
2: int_lenofcep1 ← get the length of cep1
3: int_lenofcep2 ← get the length of cep2
4: arr_intercep ← intersect(cep1, cep2)
5: int_lenofcep ← get the length of arr_intercep
6: arr_difcep1 ← difference(cep1, cep2)
7: arr_difcep2 ← difference(cep2, cep1)
8: for each l in arr_difcep1 do
9:   int_tmpsyn=0
10:  for each m in arr_difcep2 do
11:    int_tmpsyn ← synonym(l, m)
12:    if int_tmpsyn =1 then break
13:  end if
14:  if int_tmpsyn =1 then int_syn= int_syn+1
15:  end if
16:  structuralSim←(int_lenofcep+int_syn)/max(int_lenofcep1, int_lenofcep2)
17: return structuralSim
    
```

FIGURE 4 Pseudo code of process structure similarity algorithm

In the algorithm that is shown in Figure 4, intersect() and difference() are two functions that calculate

Intersection and difference set of two arrays of String type and synonym() is a function that determines whether two strings are synonyms.

For calculating the user behaviour similarity, we must use the social tagging system model. Further, as it has been discussed in the above section, a social tagging system can be represented by a triple of users, tags and resources: (U, T, R), while the resources refer in particular to the processes and activities in this paper. There is a database table named “tagging” with the columns of “user_id”, “tag_id”, “resource_id”, and “message_id”. A record “30015321, 00000271, 11120001, 30015321_00001032” represents that a user “Jim” whose id is “30015321” has bookmarked a tag “consult” whose id is “00000271” on a activity whose id is “11120001” and the tag is published in a message whose id is “30015321_00001032”. The table “tagging” can create relationships with the tables of “tags”, “users”, “process” and “message” by the id of “tag_id”, “user_id”, “resource_id” and “message_id”. The performer recommendation is used more frequently than the activity recommendation, so the user behaviour similarity algorithm is implemented by the stored procedure and the pseudo code of the user behaviour similarity algorithm is shown in Figure 5.

```

DROP PROCEDURE IF EXISTS sim_behavior;
CREATE PROCEDURE sim_behavior(
    IN p_user_id    VARCHAR(8),
    IN p_tag_id     VARCHAR(8),
    OUT p_similarity INT
)
BEGIN
    DECLARE m_tags INT DEFAULT 0;
    DECLARE m_tag INT DEFAULT 0;
    SET @COUNT_TAGS = CONCAT('SELECT
COUNT(*) INTO @TAGS_TOTAL FROM tagging where
user_id=', p_user_id);
    SET @COUNT_TAG = CONCAT('SELECT COUNT(*)
INTO @TAG_TOTAL FROM tagging where user_id=',
p_user_id, ' and tag_id= ', p_tag_id);
    PREPARE count_stmt FROM @COUNT_TAGS;
    EXECUTE count_stmt;
    DEALLOCATE PREPARE count_stmt;
    SET m_tags = @TAGS_TOTAL;
    PREPARE count_stmt FROM @COUNT_TAG;
    EXECUTE count_stmt;
    DEALLOCATE PREPARE count_stmt;
    SET m_tag = @TAG_TOTAL;
    SET p_similarity=m_tag / m_tags;
END;

```

FIGURE 5 Pseudo code of user behaviour similarity algorithm

5 Empirical Evaluation

In this section, we explain the experiment methodology that we followed to evaluate the described process recommendation method, and present the obtained results of that evaluation.

Our laboratory built a process model repository in which the process models are usually used in traditional BPM and collected from administrative examination and approval processes from administrative department of a local government (China). Firstly, we select 30 most frequently used process models as the test dataset and modify each process model by removing one or more activities and all the performers. Because the process models come from similar government administrative departments, so they can be placed in a category. After 30 minutes study of the modified dataset, 20 student volunteers can play the role of common knowledge workers that are not professional but experienced. Then the volunteers are requested to model the same business process with the similar topic on our system. The web application carrying the task of an activity is simplified by providing “accept”, “ignore”, “finish” and “recommend to others” buttons only. In our system, a social process model may be not well structured and some activities may be not strictly sequential (namely, each of the activities can be executed first) and most of performers are not clearly appointed to activities. In the experiment, the process models are put on running in turn after modelling and an activity is launched if the short URL of the web application carrying the task of the activity is sent to all users (an extreme case is that each user has the possibility to be the performer). The activities that have no strict order with each other (including implicit order) may be launched at the same time. The volunteers then are requested to select a role and accept any web application that his role is interested in or ignores it. Once a web application is accepted by a user, what the others do on it will be discarded unless the user turns to push “recommend to others” button. When the user finishes the task by pushing the “finish” button, the next activity may be launched. After all the process models are completely executed, there will be 20 social business process execution models according to the actual execution. Then five different volunteers (numbered 1, 2, 3, 4, 5) are selected to model the same business process using the commendation method after a brief introduction about the process. The modelling has a restriction on the finish time of 5 minute and the modellers can accept the recommendation or just follow their own judgments. The result is shown in Table 1.

TABLE 1 Font styles for a reference to a journal article

Volunteer Number	Time (s)	Amount of activities	Amount of similar activities with that in the sample dataset
1	207	10	7
2	93	8	8
3	312	13	9
4	109	8	8
5	158	11	9

As it can be seen from Table 1, the models that the volunteer 2 and 4 modelled have 8 activities that all exist in the sample dataset. This may be caused by the

recommendation without any modification and the similar values of total time further verify the conclusion. The other volunteers spend more time on the modelling and the activities of their models are much more. The volunteer 1 spends more time than volunteer 5 while the amount of activities in their result models raise from 10 to 11. This may be because the latter makes a better use of the recommendation result than the former. It is necessary to notice that the modelling time may be related with the user's experience and expertise. The average activities that are similar in the sample dataset exceeds 8 and this is a stronger evidence that the recommendation method is effective.

6 Conclusion and Future work

This paper presents a tag based process recommendation method for social business process modelling in the context of the social BPM. Social software is spreading quickly recently with various different application types and is affecting the modern people much more in life and work. More and more corporations have turned to develop their business or do some management by means of the tools of social software. Meanwhile, it is a challenge for traditional BPM systems that how to manage the business processes. Thus, some researchers attempt to combine the BPM and social software to realize so called social BPM. The JTangSBPMS that this

paper discussed is a new try that integrates the microblog into traditional BPM system. The tag based process recommendation method is studied for the JTangSBPMS on intelligent process modelling. The experiment results show that the method is valid and effective in the system.




Still, much work has to be carried out in the future. For example, the paper does not take into account the time factor, while a new process model may be more useful for the modeller than a process model that is modelled years ago. Furthermore, how to use the social network analysis tools to improve the process modelling still leaves much to be developed. We hope that other researchers can be inspired by this paper to pay more attention to this field, and further can propose many more comprehensive methods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant (No.61272129), National High-Tech Research Program of China(NO. 2013AA01A213), New Century Excellent Talents Program by Ministry of Education of China (No.NCET-12-0491), Zhejiang Provincial Natural Science Foundation of China (LR13F020002) and Zhejiang Provincial Natural Science Foundation of China (LY13F020027).

References

- [1] White S A, Miers D 2008 *BPMN modelling and reference guide: understanding and using BPMN* Future Strategies Inc.
- [2] Dijkman R M, Dumas M, Chun Ouyang 2008 Semantics and analysis of business process models in BPMN *Information and Software Technology* **50**(12) 1281-94
- [3] Sakr Sherif, Ahmed Awad 2010 A framework for querying graph-based business process models *Proceedings of the 19th international conference on World wide web. ACM*
- [4] Dumas Marlon, Wil M Van der Aalst, Ter Hofstede A H 2005 Process-aware information systems: bridging people and software through process technology *Wiley-Interscience*
- [5] Urban S D, et al. 2011 *The dynamics of process modelling: new directions for the use of events and rules in service-oriented computing* The evolution of conceptual modelling. Springer Berlin Heidelberg 205-24
- [6] Hermosillo G, Seinturier L, Duchien L 2010 Using complex event processing for dynamic business process adaptation *Services Computing (SCC) 2010 IEEE International Conference on. IEEE, 2010*
- [7] Loshin D 2012 *Business intelligence* Access Online via Elsevier
- [8] Erol Selim, et al. 2010 Combining BPM and social software: contradiction or chance? *Journal of software maintenance and evolution: research and practice* **22**(6-7) 449-76
- [9] Bruno G, et al. 2011 Key challenges for enabling agile BPM with social software *Journal of Software Maintenance and Evolution: Research and Practice* **23**(4) 297-326
- [10] Mathiesen P, et al. 2012 *Applying social technology to business process lifecycle management* Business Process Management Workshops. Springer Berlin Heidelberg
- [11] Van der Aalst, Wil M P 1998 The application of Petri nets to workflow management *Journal of circuits, systems, and computers* **8**(01) 21-66
- [12] Halpin H, Robu V, Shepherd H 2007 The complex dynamics of collaborative tagging *In Proceedings of the 16th international conference on World Wide Web* 211-20) ACM
- [13] Fu W T, Kannampallil T, Kang R, He J 2010 Semantic imitation in social tagging *ACM Transactions on Computer-Human Interaction (TOCHI)* **17**(3) 12
- [14] Marinho L B, Schmidt-Thieme L 2008 *Collaborative tag recommendations* In Data Analysis, Machine Learning and Applications Springer Berlin Heidelberg 533-40
- [15] Loco A 2011 *Tagging ontologies with fuzzy wordnet domains* In Fuzzy Logic and Applications Springer Berlin Heidelberg 107-14
- [16] González A, Rigau G, Castillo M 2012 *A graph-based method to improve wordnet domains* In Computational Linguistics and Intelligent Text Processing Springer Berlin Heidelberg 17-28
- [17] Hofmann T 1999 Probabilistic latent semantic analysis *In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* Morgan Kaufmann Publishers Inc. 289-96

Authors	
	<p>Yanming Ye, born on October 5, 1979, Longnan City, Gansu Province, P.R.China</p> <p>Current position, grades: Ph.D. candidate, Lecturer University studies: Artificial Intelligence Scientific interest: BPM, recommender system, social computing Publications: He has published more than 10 papers. Including: 1. "Cloud computing oriented network operating system and service platform." Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. IEEE, 2011. 2. "Social Network Supported Process Recommender System." The Scientific World Journal 2014 (2014). 3. "A Formal Modelling Method for Grid Workflow Based on Concurrent Transaction Logic." JCIT 6.2 (2011): 59-69. Experience: He is a Lecturer in Hangzhou Dianzi University and a Ph.D candidate at the College of Computer Science and Technology, Zhejiang University, China. He received his master degree in Zhejiang University, China in 2005.</p>
	<p>Yueshen Xu, born on December 11, 1989, Weifang City, Shandong Province, P.R.China</p> <p>Current position, grades: Ph.D. candidate, grade 3 University studies: Artificial Intelligence Scientific interest: service computing, recommender system, applied machine learning Publications <number or main>: He has published more than 7 papers. Including 1. "A Unified Framework of QoS-Based Web Service Recommendation with Neighbourhood-Extended Matrix Factorization." Service-Oriented Computing and Applications (SOCA), 2013 IEEE 6th International Conference on. IEEE, 2013. 2. A Highly Efficient Cloud-Based Architecture for Large-Scale STB Event Processing, 2012.7 ACM International Conference on Distributed Event-Based Systems 3. A near neighbour and maximal subgraph first based business process recommendation technique. Jisuanji Xuebao (Chinese Journal of Computers), 2013, 36(2): 263-274. 4. "Social Network Supported Process Recommender System." The Scientific World Journal 2014 (2014). Experience: He is a full-time Ph.D candidate at the College of Computer Science and Technology, Zhejiang University, China. He received his bachelor degree in XiDian University, China in 2011.</p>
	<p>Zhilin Feng, born on March 18, 1977, Hangzhou City, Zhejiang Province, P.R.China</p> <p>Current position, grades: Associate Professor University studies: Artificial Intelligence, Machine Learning Scientific interest: Data mining, Formal modelling Publications <number or main>: He has published more than 15 papers. Including: 1. "A Formal Modelling Method for Grid Workflow Based on Concurrent Transaction Logic." JCIT 6.2, 2011. 2. "Beltrami manifold denoising algorithm for ink-jet printing texture image using shape prior technology." Journal of Computational Information Systems 7.2, 2011. 3. "Distributed Workflow Service Composition Based on CTR Technology." Intelligent Computing and Information Science. Springer Berlin Heidelberg, 2011. . Experience: He received his PhD degree from school of computer science and technology in Zhejiang University in 2005.</p>

The impact of capital structure on corporate performance based on panel threshold model

Jin Wang^{1*}, Weidong Zhu²

¹*School of Management, Hefei University of Technology Hefei Anhui, China*

²*School of Management and School of Economics, Hefei University of Technology, Hefei Anhui, China*

Received 10 May 2014, www.tsi.lv

Abstract

This paper takes Value Added as an indicator of corporate performance. In considering the case of differences in growth opportunities, we use Panel Threshold Model to do an empirical analysis in the relationship between environmental management, capital structure and corporate performance of listed companies in China. The results show that: There is a weak positive correlation between environmental management and corporate performance, which means environmental management can improve corporate performance; there is a significant regime effects between capital structure and corporate performance relying on corporate growth opportunities. Capital structure and corporate performance are negatively related for low growth companies and positively related for high growth companies. This is consistent with the classical theory of capital structure, which means Value Added is more suitable as an indicator of corporate performance than profits in China.

Keywords: Panel Threshold Model, Value Added, Environmental Management, Capital Structure, Growth Opportunities

1 Introduction

As one of the important financing tools for listed companies, debt financing has a significant impact on corporate performance. Since Jensen and Meckling take Agent Theory into the framework of capital structure, debt plays more role in companies. On the one hand, debt can reduce the "Excessive Investment", so it can enhance corporate performance; on the other hand, debt will also has a negative influence on corporate performance named "Insufficient Investment". Since debt is closely related with corporate investment opportunities, and investment is limited by corporate growth opportunities, so considering the different growth opportunities between different companies to analysis the influence of debt on corporate performance is particularly necessary. In this issue, how to distinguish companies' growth division has an important influence on the conclusion of this issue. The traditional method of subjective division often leads bias to the results of regression. Therefore, In order to make a further answer to this question, this paper adopts the method Lian. et.al (2006) [1] to use Panel Threshold Model which is developed by Hansen to determine the threshold value on the data automatically.

Different scholars have different definitions of corporate performance, research has shown that the current measurement of corporate performance is not very accurate [2-5]. This paper considers the reason is measurement method. Therefore, in the measure of corporate performance we should pay more attention to the interest of other stakeholders besides shareholders.

Few scholars stand in the perspective of all stakeholders to measure corporate performance so far, and Value Added which reflects corporate value creation is a very important tool based on this perspective to evaluate corporate performance. Value Added realizes not only the interests of shareholders, but also other stakeholders to make contributions for corporate value [6], which has more broad vision than profits and has more power to encourage more stakeholders.

Environmental problems are paid more and more attention in recent years, most researches agree that environmental management is positively related to corporate performance, corporate environmental responsibility can promote corporate reputation advantages, enhance the confidence of investors, effectively use resources and market opportunities, these can also positively reflected in the capital markets; But opponents argue that, in order to improve the environment, externalities (such as the cost of dealing with pollution) was transferred into the internal of companies, this may increase the cost of operation and damage the companies' profitability. Therefore, this paper will study the relationship between environmental management and corporate performance.

The rest of this paper is as follows: based on literature review, the second part lists the research content of this paper; the third part introduces the setting of Panel Threshold Model; the forth part shows the analysis and results of empirical research; the fifth part presents the conclusion.

* *Corresponding author* e-mail: wangjin-0115@163.com

2 The research content

Value Added is the essential form of profit and fully manifestation of the socialist labour value [7]. The modern enterprise is the combination of all kinds of contract, corporate need to rely on stakeholders cooperation in order to survive. Modern corporate in the pursuit of maximizing the interests of shareholders meanwhile must consider the interests of other stakeholders, thus, Value Added which is more generalized than profits should become a target of companies [8]. The ultimate goal of corporate should not simply the pursuit of profit, it will lead to failure if corporate ignore the interests of workers and public [9].

The modern capital structure theory began in the Irrelevance Theorem proposed by Modigliani and Miller [10]. Then scholars introduce the tax effect factors, agency cost, information asymmetry to relax the strict hypotheses of MM theory, and they found that the choice of capital structure has an important influence on corporate performance [11-13]. Leland and Ross pointed out, the manager will put the debt ratio as a signal to deliver corporate performance. The evidence is that, for the companies with low market value, high debt means high bankruptcy risk and high bankruptcy cost, managers have the advantage of more information than outside investors, therefore, under no agency cost assumption, managers will choose high debt rate as far as possible. Companies with high value will try to increase debt to deliver the signal to the market, and companies with low value will try to avoid this behaviour. Therefore, debt ratio should be positively related to corporate performance [14]. Myers proposed "The Lack of Investment" base on the problem between shareholders and creditors which considers that when a company has more debt, managers will abandon the NPV is greater than zero but not enough to pay the principal and interest of the investment plan, because creditors has the priority right in claim of the cash flow. The idea is debt ratio is negatively correlated with corporate performance [15]. Jensen argues that managers usually have a tendency to grow the size of company. Therefore, there will be "Excessive Investment" [16]. This view emphasizes the conflict of interest between shareholders and managers, in order to avoid managers invest in invalid project, shareholders will force managers to use more debt to reduce free cash flow and improve corporate performance.

Integrating the perspective of Myers and Jensen, debt ratio has positive and negative two different influences on corporate performance. The fundamental reason lies in different focus. The former focus on the interest conflicts on shareholders and creditors, while the latter emphasizes the interest conflict on shareholders and managers. Stulz integrates these two kinds of relations, he argues that if managers do not hold shares in the corporation, they will increase control right through expansion of company, so they have "Excessive Investment" motivation. But in this case, shareholders will force managers to issue bonds to

reduce the "Excessive Investment", this is the positive effect of debt. But the creditor's involvement will lead the company to give up some positive NPV investment plans; this is the negative effect of debt [17]. Therefore, the positive and negative effect of debt mutual trade-off may determine the optimal capital structure to biggest corporate performance. Based on this, McConnell takes corporate growth opportunities into account; he argues that the negative impact of debt on corporate performance will be quite intense in companies with more growth opportunities. On the contrary, the positive impact of debt on corporate performance is more significant in companies with fewer growth opportunities [18]. Jung further points out, company growth opportunities will influence the optimal capital structure, and then influence corporate performance. Because with the increase of growth opportunities, consistency of the interests of managers and shareholders will be enhanced and the agency cost between them will reduce. But the agency cost between creditors and shareholders will increase with the increase of growth opportunities [19].

We argue that the different results above are mainly for the following two reasons: 1. Most scholars use the traditional financial performance to represent corporate performance, due to traditional financial indicators are in the perspective to maximize the interests of shareholders, it may lead to bias to measure corporate performance; 2. There may be a nonlinear relationship between debt ratio and corporate performance, the traditional OLS regression analysis or subjective grouping regression analysis may produce bias in the results.

Therefore, in order to obtain robust results, we use the following methods to do empirical research: 1. In order to overcome measurement bias, we stand in the perspective of all stakeholders use Value Added indicators to measure corporate performance; 2. In order to overcome estimation bias, we use Panel Threshold Model which developed by Hansen to determine the threshold value on the data automatically. In addition, due to the relationship of environmental management on corporate performance has become a focus of many scholars, this paper also introduces environmental management variables to study its effects on corporate performance.

3 The model

The literatures above show that capital structure and corporate performance may exhibit a nonlinear relationship due to different growth opportunities, which shows range effect. Because the subjective division of growth ranges may bring up estimation bias. We use Panel Threshold Model to divide growth range according to the endogenous characteristics of the data itself, and then study the relationship between capital structure and corporate performance in different growth range. The model setting and estimation method are as follows:

3.1 SINGLE THRESHOLD MODEL

Single Threshold Model sets as follows:

$$Performance_{it} = u_i + \theta x_{it} + \beta_1 Lev_{it} I(grow_{it} \leq \gamma) + \beta_2 Lev_{it} I(grow_{it} \geq \gamma) + \varepsilon_{it} \tag{1}$$

where i is company, t is year, $Performance_{it}$ and Lev_{it} respectively represent corporate performance and capital structure. x_{it} is a group of control variables which influence corporate performance, including environmental management, company size, asset structure, the liquidity of shares and profitability. θ is the corresponding coefficient vector. $grow_{it}$ is the threshold variable, in this paper it is corporate growth opportunities, γ is a particular threshold value. $I(\cdot)$ is an index function. u_i reflects companies' individual effects, which are the unobservable factors such as corporate culture, management ability and leadership qualities etc. $\varepsilon_{it} \sim i.i.d. N(0, \sigma^2)$ is random disturbance. In order to estimate the values of parameters, we need each observation minus the average value within group to eliminate the individual effect u_i , e.g. $Performance_{it}^* = Performance_{it} - \frac{1}{T} \sum_{t=1}^T Performance_{it}$

the transformed model is:

$$Performance_{it}^* = \theta x_{it}^* + \beta_1 Lev_{it}^* I(grow_{it} \leq \gamma) + \beta_2 Lev_{it}^* I(grow_{it} \geq \gamma) + \varepsilon_{it}^* \tag{2}$$

Then we stack all observation and use the matrix form to express (2) as:

$$Performance_{it}^* = X^*(\gamma)\beta + \varepsilon^* \tag{3}$$

For threshold value γ , we can use OLS regression to estimate (3) to obtain estimated value of β :

$$\hat{\beta}(\gamma) = (X^*(\gamma)' X^*(\gamma))^{-1} X^*(\gamma)' Performance_{it}^* \tag{4}$$

Corresponding sum of squared residuals is:

$$S_1(\gamma) = \hat{e}^*(\gamma)' \hat{e}^*(\gamma) \tag{5}$$

where $\hat{e}^*(\gamma) = Performance_{it}^* - X^*(\gamma)\hat{\beta}(\gamma)$ is residual vector.

We can minimization $S_1(\gamma)$ in (5) to obtain the estimated value of γ , i.e.

$$\hat{\gamma} = \arg \min_{\gamma} S_1(\gamma) \tag{6}$$

Once we get $\hat{\gamma}$, we can then get $\hat{\beta} = \hat{\beta}(\hat{\gamma})$, residual

vector $\hat{e}^* = \hat{e}^*(\hat{\gamma})$ and the square of residuals $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\gamma}) = \frac{1}{n(T-1)} \hat{e}^* \hat{e}^* = \frac{1}{n(T-1)} S_1(\hat{\sigma})$, n is the number of companies.

We should do two hypotheses testing after we obtain the parameter estimated values. One is whether the threshold effect is significant; the other one is whether the threshold estimated value is equal to its real value. The null hypothesis of first test is $H_0: \beta_1 = \beta_2$, corresponding alternative hypothesis is $H_1: \beta_1 \neq \beta_2$, the test statistic is:

$$F_1 = \frac{S_0 - S_1(\hat{\gamma})}{\hat{\sigma}^2} = \frac{S_0 - S_1(\hat{\gamma})}{S_1(\hat{\gamma}) / n(T-1)} \tag{7}$$

where S_0 is sum of squared residuals under null hypothesis H_0 . Under null hypothesis H_0 , the value of threshold γ is unrecognized. Therefore, the distribution of F_1 is not standard. Hansen (1999) shows that "Bootstrap" can obtain its asymptotic distribution; the p value based on this structure is also asymptotically valid [20]. The null hypothesis of second test is $H_0: \hat{\gamma} = \gamma_0$, corresponding likelihood ratio statistic is:

$$LR_1(\gamma) = \frac{S_1(\gamma) - S_1(\hat{\gamma})}{\hat{\sigma}^2} \tag{8}$$

The distribution of this statistic is also not standard; Hansen provides a simple formula to calculate the non rejection region. That is we cannot reject the null hypothesis when $LR_1(\gamma_0) \leq c(\alpha)$, where $c(\alpha) = -2 \ln(1 - \sqrt{1 - \alpha})$, α is significant level.

2.2 MULTIPLE THRESHOLD MODEL

There is only one threshold in Model (1), but there may be more than one threshold in many cases. E.g. Double Threshold Model is setting as follows:

$$Performance_{it} = \beta_1 Lev_{it} I(grow_{it} \leq \gamma_1) + \beta_2 Lev_{it} I(\gamma_1 \leq grow_{it} \leq \gamma_2) + \beta_3 Lev_{it} I(grow_{it} > \gamma_2) + u_i + \theta x_{it} + \varepsilon_{it} \tag{9}$$

where $\gamma_1 < \gamma_2$. Here we only focus on Double Threshold Model, because it can be extended easily to the case of Multiple Threshold Model. In order to reduce the computation, we use "Circulation Method" to estimate Model (9). In a model with structure mutation, this method can obtain the consistent estimation of parameters, such as Lian et al. (2006). The first step, let $S_1(\gamma)$ be sum of squared residuals in Single Threshold Model defined by (5), γ_1 is the estimated value of threshold when $S_1(\gamma)$ is minimum. Bai (1997) shows that

both for γ_1 and γ_2 , $\hat{\gamma}_1$ is the consistent estimation of γ_1 .

Fixed $\hat{\gamma}_1$ obtained in the first step to estimate Model (9), the screening criteria for the second step

$$is S_2^{\gamma}(\gamma_2) = \begin{cases} S(\hat{\gamma}_1, \gamma_2) & \text{if } \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & \text{if } \gamma_2 < \hat{\gamma}_1 \end{cases}, \text{ and the estimation of threshold}$$

in the second step is $\hat{\gamma}_2^{\gamma} = \arg \min_{\gamma_2} S_2^{\gamma}(\gamma_2)$.

Bai shows that $\hat{\gamma}_2^{\gamma}$ is asymptotically efficient, but the estimation of $\hat{\gamma}_1$ does not have the property [21]. This is

because when we estimate $\hat{\gamma}_1$, sum of squared residuals contains the interval we ignored. But due to $\hat{\gamma}_2^{\gamma}$ is

asymptotically efficient, we can fix $\hat{\gamma}_2^{\gamma}$, and then re-estimation, at this time, the screening criteria is

$$S_1^{\gamma}(\gamma_1) = \begin{cases} S(\gamma_1, \hat{\gamma}_2^{\gamma}) & \text{if } \gamma_1 < \hat{\gamma}_2^{\gamma} \\ S(\hat{\gamma}_2^{\gamma}, \gamma_1) & \text{if } \hat{\gamma}_2^{\gamma} < \gamma_1 \end{cases}, \text{ final we get the optimized}$$

consistent estimation $\hat{\gamma}_1 = \arg \min_{\gamma_1} S_1^{\gamma}(\gamma_1)$. Hypothesis Test in Double Threshold Model is similar to Single Threshold Model, we do not repeat it here.

4 Empirical analyses

4.1 SAMPLES AND PROXY VARIABLES

Our data is obtained from CSMAR developed by Shenzhen GTA Information Technology Company. In this paper we select Shanghai and Shenzhen A-share listed companies as samples, studying period is 2003-2011. We screen the data according to the following steps: (1) remove Financial Companies; (2) remove companies whose asset-liability ratio exceeds 100% (3) remove companies whose growth rate of total assets surpasses 150% for there may exist merger behaviour in these companies; (4) the key financial variables are winsored at 1st and 99th percentiles to avoid the influence of outliers. Ultimately, we obtain 1002 companies and 9018 observations.

Table 1 lists the definition and descriptive statistics of proxy variables in Model (1), considering the profit index is easy to control, we use Value Added index as corporate performance. We adopt ‘‘add algorithm’’ to calculate Value Added [6], the calculation method is:

$$Value\ Added = Employees\ Income + Creditors\ Income + Shareholder\ Income + Governmenr\ Income + Corporate\ Income \quad (10)$$

Due to the data of environmental management is not easy to get, in this paper the proxy variable of environmental management is taken from the companies’ annual report and financial statements. If they mention of the behaviour of environmental management such as environmental governance, environmental protection, environmental technology etc., the environmental management variable is 1, otherwise 0.

TABLE 1 Descriptive statistics of samples

Variable	Variable meaning	Calculation method	Maximum	Minimum	Mean	SD
Performance	Corporate performance	Value Added/ Total assets	0.314	-0.023	0.125	0.083
Environment	Environmental management	Manage the environment is 1, otherwise 0	1.000	0.000	0.037	0.188
Lev	Capital structure	Total liabilities/ Total assets	0.857	0.189	0.526	0.181
Lnasset	Company size	LN(total assets)	23.814	19.874	21.598	1.048
Tang	Asset structure	(Fixed assets + Inventories)/ Total assets	0.756	0.158	0.467	0.168
Tshr	Share liquidity	Shares outstanding/ Total share capital	1.000	0.273	0.608	0.247
Prof	Profitability	Net profit/ Main business revenue	0.308	-0.234	0.057	0.113
Grow	Growth opportunities	Growth rate of total assets	0.580	-0.189	0.114	0.188

4.2 THE EMPIRICAL RESULTS

In order to determine the form of model we must determine the number of threshold. We successively estimate Model (9) under no threshold, one threshold,

two thresholds and three thresholds, the F statistics and the Bootstrap P values are shown in Table 2. As shown in Table 2, the single threshold and double threshold effect is very significant, but the triple threshold effect is not significant, so we only analyse Double Threshold Model.

TABLE 2 The threshold effect test

	F value	P value	BS times	Critical value		
				1%	5%	10%
Single Threshold	47.643***	0.000	500	7.813	5.128	3.381
Double Threshold	11.061**	0.012	500	11.281	7.397	5.350
Triple Threshold	2.517	0.132	500	8.430	4.488	3.013

Note: ***, ** and * respectively at 1%, 5% and 10% significant level, same below.

Two threshold values of the model as well as the estimated 95% confidence intervals are shown in Table 3. With the likelihood ratio function drawn in Figure 1 and Figure 2, we can clearly see the constructing process of the estimation values and confidence intervals of threshold. The estimated value of the threshold refers the values of γ when the likelihood ratio test statistic LR is 0. In our Double Threshold Model are respectively 0.215 (Figure 1) and 0.560 (Figure 2). 95% confidence interval for each threshold value is the interval of γ for all LR less than the threshold level of 7.35, which is significantly at 5% level (corresponding to the broken line in Figure).

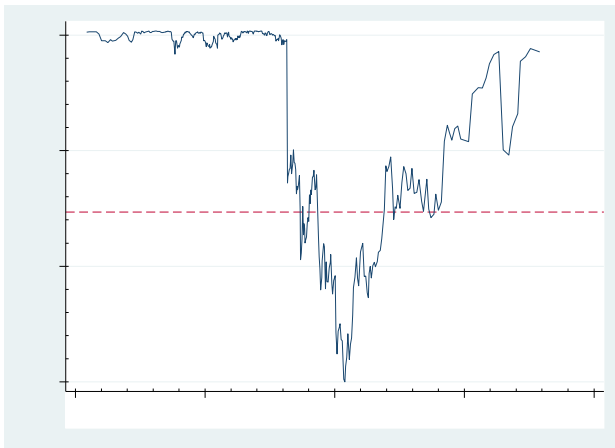


FIGURE 1 Estimation and Confidence Intervals for the First Threshold

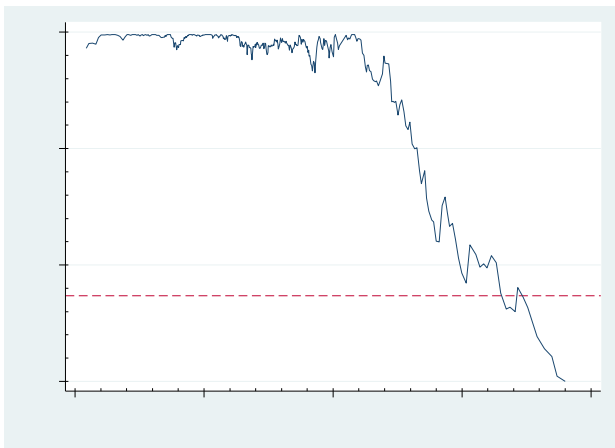


FIGURE 2 Estimation and confidence intervals for the second threshold

TABLE 3 Threshold estimation results

	Estimated value	95% Confidence interval
Threshold Value $\hat{\gamma}_1$	0.215	[0.147, 0.353]
Threshold Value $\hat{\gamma}_2$	0.560	[0.461, 0.560]

We can list companies in low growth ($Grow \leq 0.215$), moderate growth ($0.215 < Grow \leq 0.560$) and high growth ($Grow > 0.560$) three types according the two threshold values. And respectively do regression analysis according to the different growth interval. The parameter estimation,

t value under constant variance (t_{OLS}) and t value under heteroscedastic variance (t_{White}) is in TABLE 4. As shown in Table 4, the result of four control variables used in this paper is: company size and corporate performance is negatively correlated, asset structure, share liquidity and profitability are positively related to corporate performance, this is consistent with previous scholars.

TABLE 4 The result of regression

Variable	Coefficient	t_{OLS}	t_{White}
Environment	0.006	0.99	0.68
Lnasset	-0.015	-12.61***	-7.39***
Tang	0.021	4.35***	3.05***
Tshr	0.036	14.77***	9.78***
Prof	0.439	76.22***	40.17***
Lev_a	0.009	3.72***	3.03***
Lev_b	-0.007	-2.69***	-1.76*
Lev_c	-0.025	-5.59***	3.03***
CONS	0.389	15.62***	9.19***

Note: (1) Lev_a, Lev_b and Lev_c is the debt ratio in low, moderate and high growth interval respectively; (2) t_{OLS} is t value under constant variance, t_{White} is t value under heteroscedastic variance; (3) R^2 of this model is 0.435, p value of F test is 0.000.

Our focus is the relationship between environmental management, capital structure and corporate performance. We find that the environment variable has positive effect on corporate performance, but not significant, which may be because the measurement of environmental management is not accurate. As Table 4 shows, the debt ratio is positively correlated with corporate performance in low growth companies, and the significant level is 1%; but for the moderate growth and high growth companies, debt ratio and corporate performance is significantly positive correlation. Our results is just opposite with Lian et al. (2006)'s, but consistent with McConnell and Jung's. This proves that Value Added is more suitable for corporate performance than traditional financial index. Classical capital structure theory could not be confirmed when taking profit as proxy variable of corporate performance, but the theory is confirmed when taking Value Added as proxy variable of corporate performance.

5 Conclusions

5.1 ENVIRONMENTAL MANAGEMENT CAN IMPROVE CORPORATE PERFORMANCE

As shown in Table 4, environmental management has a positive effect on corporate performance, although this effect is not significant. We argue that the reason of not significant may be the measurement of environmental management. Due to the data we extracted is according to companies' annual report and financial statements, we will not be able to collect the data if many companies conducted in environmental management work, but its annual report and financial statements were not mentioned. This may cause bias on measurement of environmental management, which lead coefficient is not significant. Recently environmental issues have become a major problem affecting corporate performance; the

leader of company should not only pay attention to the company's operating performance, but also should pay attention to environment protection, ecology, etc. Because it is also possible to improve corporate performance when the environment is protected.

5.2 THE IMPACT OF CAPITAL STRUCTURE ON CORPORATE PERFORMANCE CHANGES ALONG WITH THE GROWTH OF COMPANIES

As the results shown in TABLE 4, there is a significant positive correlation between debt ratio and corporate performance in low growth companies, but there is a significant negative correlation between debt ratio and corporate performance in moderate growth and high growth companies. This is consistent with the view proposed by McConnell: "The negative impact of debt on

corporate performance will be quite intense in companies with more growth opportunities. On the contrary, the positive impact of debt on corporate performance is more significant in companies with fewer growth opportunities." This reflects that Value Added is more suitable for corporate performance than profits in China, because the result is consistent with classical capital structure theory when we use Value Added index to measure corporate performance.

Acknowledgments

This paper is supported by Doctoral Fund of the Ministry of Education funded project *Value Added distribution theory and empirical research in China's listed companies based on the perspective of stakeholders.* (20100111110015)

References

- [1] Lian Y J, Cheng J 2006 Relationship between capital structure and performance with different growth opportunities *Modern Economic Science* 28(1) 97-103+128 (in Chinese)
- [2] Wang Y T, Wang L L, Yang Y 2010 Ownership nature of ultimate controller, debt-related tax shields and capital structure *Economic Research Journal* (9) 122-36 (in Chinese)
- [3] Wang X, Wang L 2012 A research on financial contingency theory of capital structure adjustment-the case study of J group *Accounting Research* (10) 65-70+96 (in Chinese)
- [4] Sheng M Q, Zhang M, Ma L J, Li H 2012 State-owned right, soft budget constraints and the dynamic adjustment of capital structure *Management Word* (3) 151-7 (in Chinese)
- [5] Wu Z X, Zhang Y, Zhang W 2013 Credit policy and capital structure of corporation- evidence from listed firms in China *Accounting Research* (3) 51-8+96 (in Chinese)
- [6] Zhu W D, Wang J 2013 The Relative Relationship Between Employees Income Rate and Enterprise Value Based on Value Added *Metalurgia International* 18(11) 49-52
- [7] Huo D F 1991 New theory of Value Added *Economic Review* (1) 33-5 (in Chinese)
- [8] Zhao L P 2002 Further discussion on enterprise Value Added Table *Public Finance Research* (1) 61-5 (in Chinese)
- [9] Du X Q 1996 Statement of Value Added *Communication of Finance and Accounting* (10) 6-7 (in Chinese)
- [10] Modigliani F, Miller M H 1958 The Cost of Capital, Corporation Finance and the Theory of Investment *American Economic Review* 48(3) 261-97
- [11] Jiang L, Song C, Wei C 2013 Study on the business cyclical fluctuation and the adjustment method of capital structure *Accounting Research* (7) 28-34+96 (in Chinese)
- [12] Li M X 2011 Study on the relationship between capital structure and China's commercial banks 'in three principles' *Management World* (2) 173-4+7 (in Chinese)
- [13] Wang Z W, Wang S M, Zhu W X 2011 Regulation on equity financing and the optimal capital structure *Management World* (2) 40-8+187 (in Chinese)
- [14] Ross S A 1977 The determination of financial structure: the incentive signaling approach *Bell Journal of Economics and Management Science* 8(1) 23-40
- [15] Myers S 1977 Determinants of Corporate Borrowing *Journal of Financial Economics* 5(2) 147-75
- [16] Jensen M C 1986 Agency Cost of Free Cash Flow, Corporate Finance, and Takeovers *American Economic Review* (76) 323-9
- [17] Stulz R 1990 Management Discretion and Optimal Financing Policies *Journal of Financial Economics* 26(1) 3-27
- [18] McConnell J, Servaes H 1995 Equity ownership and the two faces of debt *Journal of Financial Economics* 39(1)
- [19] Jung K, Kim Y C, Stulz R M 1996 Timing, investment opportunities, managerial discretion, and the security issue decision *Journal of Financial Economics* (42) 159-85
- [20] Hansen B W 1999 Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference *American Economic Review* (93) 345-68
- [21] Bai J 1997 Estimating multiple breaks one at a time *Econometric Theory* (13) 315-52

Authors



Jin Wang, born on January 15, 1985, Shenyang Liaoning China

Current position, grades: Doctoral student of School of Management

University studies: Hefei University of Technology

Scientific interest: Finance and Accounting, Value Added Accounting, and Decision Making Evaluation

Publications: 2

Experience: He received master degree of Accounting at Hefei University of Technology and now he is a doctoral student of School of Management, Hefei University of Technology.



Weidong Zhu, born on January 25, 1962, Xianju Zhejiang China

Current position, grades: professor at School of Economics

University studies: Hefei University of Technology

Scientific interest: Artificial Intelligence, Decision Support Systems, Accounting Information, and Decision Making.

Publications: Very Much

Experience: He received the Ph.D. degree in management science and engineering from Hefei University of Technology. Currently, he is a professor at Hefei University of Technology, School of Economics. And he is a doctoral supervisor in School of Management, Hefei University of Technology.

Short-term prediction of wind power based on self-adaptive niche particle swarm optimization

Hong Zhang^{1, 2*}

¹ Department of Electrical Engineering, Southeast University, Nanjing, Jiangsu 210096 China

² Key Laboratory of Technical and Device of Smart Grid, Nanjing, Jiangsu 210096 China

Received 12 May 2014, www.tsi.lv

Abstract

Connecting wind power to the power grid has recently become more common. To better manage and use wind power, its strength must be predicted precisely, which is of great safety and economic significance. Speed sensors are widely applied, it make prediction of wind power more accurate. In this paper, the short-term prediction of wind power is based on self-adaptive niche particle swarm optimization (NPSO) in a neural net. Improved PSO adopts the rules of classification and elimination of a niche using a self-adaptive nonlinear mutation operator. Compared with the traditional method of maximum gradient, NPSO can skip a local optimal solution and approach the global optimal solution more easily in practice. Compared with the basic PSO, the number of iterations is reduced when the global optimal solution is obtained. The method proposed in this paper is experimentally shown to be capable of efficient prediction and useful for short-term power prediction.

Keywords: Speed sensor, PSO, Niche, Short-term power prediction, Neural net

1 Introduction

Wind power is a renewable energy source that is becoming increasingly popular for application in the grid because of its environmentally friendly and low-cost properties. However, because the power fluctuates with the wind strength, connecting wind power to the grid is challenging. To make the use of wind power reasonable and reduce its negative effects on the power grid, scientists in many countries have been working to develop methods to predict the power of the wind generators, which is of great importance to the economical distribution and operation of the power grid. Advanced wind speed sensor makes it possible to accurately predict short-term wind power and plays an important role in the wind power prediction. Denmark was among the first countries to develop a system of power prediction for wind power [1]. Prediktor is the wind power work prediction system developed by Ris National Laboratory of Denmark, which mainly applies physical models [2]. ANEMOS, a research project sponsored by the European Union, combines physical and statistical methods [3, 4]. The eWind is a system developed by AWS Truewind in America [5]. The highly precise mathematical models of atmospheric physics and adaptive statistical models are combined; the velocity of the wind and the power of the wind power plants have been investigated in studies based on time serials and neural networks [6-8]. At present, quite some few PSO alternatives such as Wang et al (2012) [9], Pousinho et al

(2010) [10] or Pratheepraj et al (2011) [11] for short-term wind power prediction, and so on.

The back propagation (BP) neural network is the mostly widely used neural network. The classic BP learning law is typically used in BP neural networks to determine net-work connection weights. However, this technique is slow in practice and may lead to a local optimal solution. In this paper, the short-term power prediction of the wind power is based oneself-adaptive niche particle swarm optimization (NPSO) in a neural network. Improved PSO adopts the rules of classification and elimination of a niche and uses a self-adaptive nonlinear mutation operator. Compared with the traditional method of maximum gradient, NPSO can skip a local optimal solution and approach the global optimal solution more easily in practice. Compared with the basic PSO, the number of iterations is reduced when the global optimal solution is obtained. The method proposed in this paper is experimentally shown to be capable of efficient prediction and useful for short-term power prediction.

2 The principle of prediction of wind power based on speed data from mechanical sensors

In 1926, Betz proposed general theory of Betz about aerodynamic action of the wind. From this theory, we get the formula related to the power output of wind turbines and wind speed as shown in Equation 1.

$$P_1 = 8/27 \rho S V^3 C_P, \quad (1)$$

* Corresponding author e-mail: hazh0216@163.com

ρ is the density of air, S is cross-sectional area of airflow through wind turbine fan. V is the average wind speed through the fan. C_p is the actual utilization of wind turbine efficiency.

$$n=2NV/D \tag{2}$$

$$N=\frac{1}{2}\rho A v_m^3 \tag{3}$$

$$D=2\pi r^2 A b_m \tag{4}$$

The wind speed measured by cup anemometer is based on the formula 2-4. In Equations (2), (3) and (4), n is the rotational speed of cup anemometer; A is the cross sectional area of the cups; a_m and b_m are intrinsic parameters of cup anemometer. From the formula (1)-(4), we can obtain the wind turbine power from the wind speed measured by wind speed sensor. When we want to know the wind power of sometime later in the future, we can use hybrid swarm intelligence theory based on speed data from speed sensors. It is described the prediction of wind power as shown Figure 1

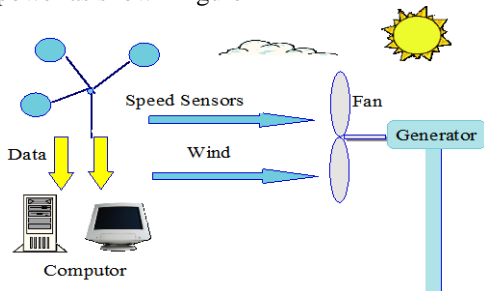


FIGURE 1 Schematic diagram for the prediction of wind power

3 Theoretical basis for improved self-adaptive PSO

3.1 THEORETICAL BASIS FOR BASIC PARTICLE SWARM OPTIMIZATION

In 1995, J. Kennedy and R. C. Eberhart developed PSO [12, 13], which aims to simulate a simple social system, such as a bird flock searching for foods, to study and explain complex social behaviour. In basic PSO, every candidate solution is compared to a bird searching the space and is called a particle. The position and velocity of a particle is denoted as $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ and $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, respectively. At the initial stage, a swarm of particles is randomly selected. Then, the swarm is updated according to the best known positions of individual particles and the entire swarm. The equations defining the position and velocity of the particles are shown below:

$$\begin{aligned} v_{id}(k+1) &= wv_{id}(k) + c_1r_1(p_{id}(k) - x_{id}(k)) \\ &\quad + c_2r_2(g_{id}(k) - x_{id}(k)) \\ x_{id}(k+1) &= x_{id}(k) + v_{id}(k) \end{aligned} \tag{5}$$

In Equations (5) and (6), p is the best known position of a particle and g is the best known position of the entire swarm; $i = 1, 2, \dots, n$; D is the dimension of a particle; k is the k -th iteration; d is the d -th dimension; k_{max} is the maximum number of iterations; w is the inertia weight; w_{ini} is the initial inertia weight; w_{end} is the final inertia weight; c_1 and c_2 are learning factors; and r_1 and r_2 are uniform random numbers in the range $[0, 1]$.

3.2 ADAPTIVE NICHING PARTICLE SWARM OPTIMIZATION

Basic PSO may lead to premature convergence to a local optimum, thus affecting the quality of the solution. The probability of prematurity can be reduced by mixing basic PSO with other algorithms or by adopting a comprehensive strategy. Niche technology simulates ecological balance, i.e., a species evolves to establish a surviving niche in a larger environment, which reflects the evolutionary rule of survival of the fittest. Goldberg and Richardson described niche technology based on a sharing mechanism in [14], and Brits et al., described NPSO in [15, 16]. The following formulae are based on adaptive NPSO:

$$\begin{aligned} v_{id}(k+1) &= wv_{id}(k) + c_1r_1(p_{id}(k) - x_{id}(k)) + \\ &\quad c_2r_2(g_{id}(k) - x_{id}(k)) + c_3r_3(\bar{p}_{id}(k) - x_{id}(k)) \end{aligned} \tag{7}$$

$$x_{id}(k+1) = x_{id}(k) + v_{id}(k)$$

$$w = (w_{ini} - w_{end}) \exp(-1/[1 + (1 + \frac{k}{k_{max}})]) + w_{end} \tag{8}$$

In Equations (7) and (8), \bar{p}_{id} is the best known position of a sub-swarm; c_3 is the learning factor; and r_3 is a uniform random sequence in the range $[0, 1]$.

The diversity selection of the swarm regulates the adaptability of individual particles by reflecting the sharing functions among them, upon which the later evolutionary process is selected, to create an evolved environment and to realize swarm diversity.

The adaptive mutation operator adopts an adaptive non-linear decreasing inertia weight function [17]. The decreasing velocity of the inertia weight is accelerated in the first iteration of the algorithm to achieve a more efficient solution.

3.3 THE MAIN STEPS OF THE IMPROVED PSO ALGORITHM

The main steps of the improved PSO algorithm are as follows:

- Step 1 Start.
- Step 2 Generate the initial population by chaotic iteration.
- Step 3 Initialize parameters.

- Step 4 Select a particle randomly and divide all of the particles evenly into m small niche subpopulation based on adaptive functions.
- Step 5 Establish the initial velocity of the particles randomly.
- Step 6 Set the initial position of the present particle as the individual historical optimal value, pbx; set the historical optimal value of the optimal individual in each subpopulation as the population historical optimal value, $\bar{p}bx$; and set the historical optimal value of all of the particles as the overall historical optimal value, gbx.
- Step 7 When k is less than the maximum number of iterations, the following cycle of operations is performed for each subpopulation:
 - I) Calculate the inertia weight, threshold value, and calibration coefficient.
 - II) Update the velocity and position of every particle within each subpopulation.
- Step 8 Adopt a niche elimination strategy.
- Step 9 Determine whether the convergence conditions are met; if so, stop the calculation and output the results; if not, go to Step 6.
- Step 10 End.

3.4 TESTING THE IMPROVED PSO ALGORITHM USING STANDARD TEST FNCTIONS

To test the performance of the improved PSO algorithm, two standard testing functions are selected: the 2-D Rosenbrock function and 2-D Rastrigin function. Standard testing functions are commonly employed in the optimization literature to evaluate the efficiency of new algorithms [18, 19]. The two standard testing functions have numerous local optima and a global minimum that is very difficult to locate.

3.4.1 The 2-D Rosenbrock Function

The 2-D Rosenbrock function is given by Equation (9):

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{9}$$

For the 2-D Rosenbrock function in this paper, the global minimum is $f_{global} = 0$ as $x = (1, 1)$, but the valley in which the minimum lies has steep edges and a narrow ridge. The tip of ridge is also steep. Figure 2 illustrates the main characteristics of the 2-D Rosenbrock function.

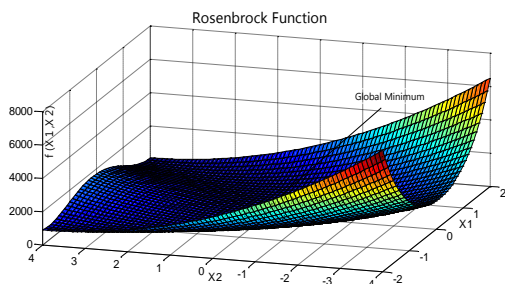


FIGURE 2 Graph of the Rosenbrock function

3.4.2 The 2-D Rastrigin function

The 2-D Rastrigin function is given by Equation (10):

$$g(x_1, x_2) = x_1^2 + x_2^2 - 10[\cos(2\pi x_1) + \cos(2\pi x_2)] + 20 \tag{10}$$

For the 2-D Rastrigin function employed in this paper, the global minimum is $f_{global} = 0$ when $x = (0, 0)$. There are many local minima arranged in a lattice configuration, as shown in Figure 3. Figure 3 illustrates the main characteristics of the 2-D Rosenbrock function.

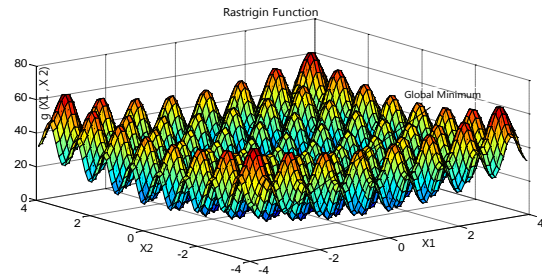


FIGURE 3 Graph of the Rastrigin function.

The global minima of the 2-D Rosenbrock function and 2-D Rastrigin function can be located by simulation computation based on the improved PSO algorithm. Thus, the model based on the improved PSO can be used in practice.

4 Neural Network Model Based on Self-Adaptive Niche PSO

4.1 THERORETICAL BASIS FOR THE BASIC NEURAL NETWORK

Since the insightful study of the neural network in the 1980s [20, 21], neural networks have been widely applied to the industrial field. The artificial intelligence neural network is a complex nonlinear system. The artificial neural network is also a nonlinear mapping system with good self-adaptability and can be used to identify any complicated state or process.

Figure 4 describes a simple artificial intelligence neural network. The basic principle of the neural network model to process information is that the input signal $X(i)$ acts on the intermediate node (the hidden layer), leading to a result from the output node, which utilizes a non-linear transformation and generates an output signal $Y(k)$ by adjusting $W(ij)$, the value relating to the input nodes and hidden layer nodes. $T(jk)$, the value relating to the hidden layer nodes, the output node, and their respective values, is reduced by repetitive learning training; the network parameters (weights and threshold values) relating to the minimum error are determined. The training continues until the error reaches the threshold value. The BP neural network model is expressed in Equation (11):

$$\begin{aligned}
 O_j &= f(\sum W_{ij} \times X_i - q_j) \\
 Y_k &= f(\sum T_{jk} \times O_j - q_k)
 \end{aligned}
 \tag{11}$$

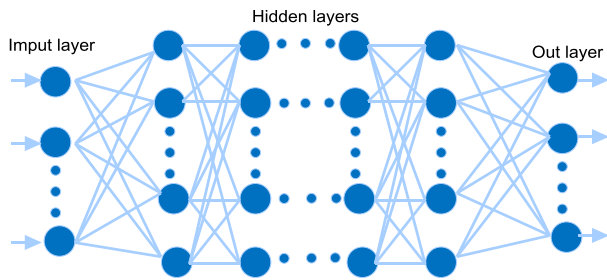


FIGURE 4 Artificial intelligence neural network

4.2 THE STEPS OF THE PREDICTION ALGORITHM BASED ON SELF-ADAPTIVE NPSO NEURAL NETWORK

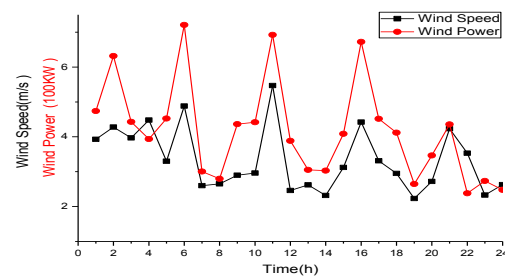
The main steps of prediction algorithm based on the self-adaptive NPSO neural network are as follows:

- Step 1 Start.
- Step 2 Input the initial values and target values of the samples.
- Step 3 Initialize the coupling weight values and thresholds.
- Step 4 Convert connection weights and thresholds to particles.
- Step 5 Divide the initial population into several small niche subpopulations.
- Step 6 Calculate the adaptive values of the particle swarm.
- Step 7 Determine the best known positions of the individuals, sub-populations, and overall population.
- Step 8 Adjust the adaptability and inertia weight and update the velocity and position of the particles.
- Step 9 Judge whether the niche update conditions are met. If not, go to Step 6.
- Step 10 Run the niche optimization rules.
- Step 11 Judge whether the maximum time is reached. If not, go to Step 6.
- Step 12 Determine the coupling value and threshold.
- Step 13 End.

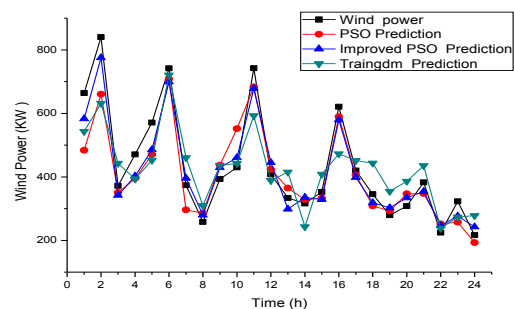
5 Predictive Analysis of the Neural Network Based on Self-Adaptive NPSO

The power prediction model is established by the neural network based on self-adaptive NPSO (improved PSO). The power of a wind generator in Dongtai (Jiangsu, China) was predicted in 2008 based on the meteorological data and data for the power generated by the wind generator in the previous months. The predictive models for the neural network are based on PSO, NPSO, and Traingdm. First, the original data related to wind speed and wind power must be processed and normalized by advanced mathematical methods [22, 23]. For example, the model will observably decrease systematic

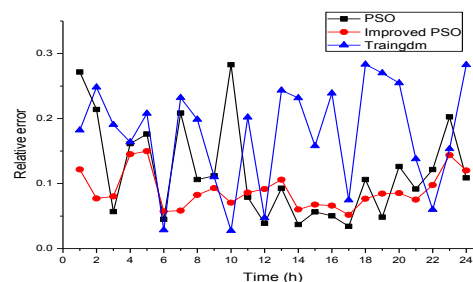
error when the origin data have been processed by the Kalman filter described in the literature [24, 25]. All predictive models are trained beforehand. Figure 5 illustrates the main characteristics obtained from different prediction models 3 h ahead. Figure 5(a) illustrates that higher wind powers generally correspond to higher wind speeds. Figure 5(b) presents the measured power and forecasted power based on PSO, improved PSO, and Traingdm. Comparing the results of the three methods, the forecasted wind power curve based on the improved PSO is the closest to the measured power in Figure 5(b). Figure 5(c) presents the relative error from different predictions. The minimum relative error of the forecast wind power is obtained by the improved PSO method. Figure 5(d) illustrates the frequency and probability from different prediction models based on PSO, improved PSO, and Traingdm. The probability of a relative error of less than 0.1 for the improved PSO method is greater than those of PSO and Traingdm. Thus, the prediction accuracy of the improved PSO method is better than those of PSO and Traingdm. The absolute error, relative error, means absolute error, mean relative error, standard deviation, and relative standard



(a)



(b)



(c)

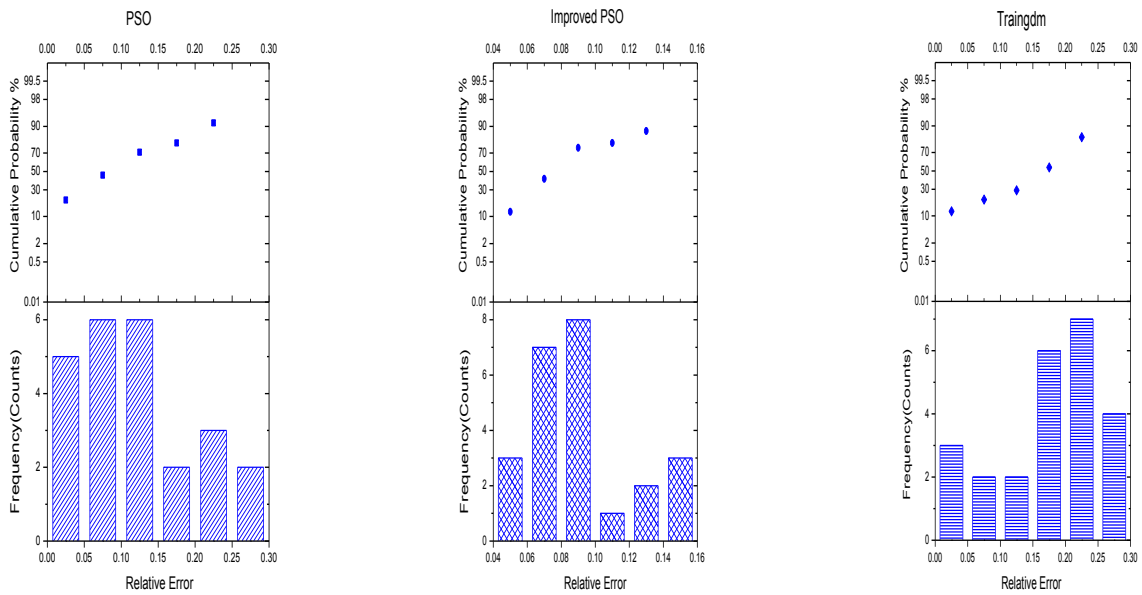


FIGURE 5 Main characteristics obtained from the three different predictions

(a) Wind speed and wind power. (b) The measured power and forecasted power based on PSO, improved PSO, and Traingdm. (c) Relative error from different prediction models based on PSO, improved PSO, and Traingdm. (d) Frequency and probability from different prediction models based on PSO, improved PSO, and Traingdm.

Deviation and interval probability in this paper are illustrated by Equation (12) [26]. From this data of Figure 5(d), we can determine the best method of the three based on PSO, improved PSO, and Traingdm.

$$\begin{aligned}
 \text{absolute error} &= | \text{forecast} (i) - \text{measure} (i) | \\
 \text{mean absolute error} &= \frac{ | \text{forecast} (i) - \text{measure} (i) | }{ n } \\
 \text{relative error} &= \frac{ | \text{forecast} (i) - \text{measure} (i) | }{ \text{measure} (i) } \\
 \text{mean relative error} &= \frac{ \text{relative error} }{ n } \\
 \text{standard deviation} &= \sqrt{ \frac{ \sum_{i=1}^n (\text{forecast} (i) - \frac{1}{n} \sum_{i=1}^n \text{forecast} (i))^2 }{ n-1 } } \\
 \text{relative standard deviation} &= \frac{ \text{standard deviation} }{ \frac{1}{n} \sum_{i=1}^n \text{forecast} (i) } \\
 \text{interval probability} &= \frac{ \text{frequency}(\text{counts}) }{ n }
 \end{aligned}
 \tag{12}$$

6 Conclusions

In this paper, a predictive model for neural networks based on self-adaptive NPSO is established. Using model analysis, experiments, and comparison with predictive models based on other algorithms, the model is shown to be more precise than the other two models considered; furthermore, it has the lowest absolute variance, demonstrating its effectiveness. The reliability of the model is significantly related with the precision of the weather forecast, but by combining with the weather predictive model, the dependence will be lowered. With computers becoming increasingly powerful, the predictive method of the neural network based on hybrid multi-algorithms will be useful in the future.

Acknowledgments

The author gratefully acknowledges Prof. Xueliang Huang School of Electrical Engineering, Southeast University. This paper is partly supported by the National High Technology Research and Development Program (863 Program) (2011AA05A107).

References

[1] Landberg L, Watson S J 1994 Short-term prediction of local wind conditions *Bound.-Lay. Meteorol* **70** 171-95
 [2] Landberg L, Giebel G, Nielsen H A, et al. 2003 Short-term prediction-an overview *Wind Energy* **6** 273-80
 [3] Martí I, Kariniotakis G, Pinson P, et al. 2006 Evaluation of Advanced Wind Power Forecasting Models—Results of the Anemos Project *In Proceedings of the European Wind Energy Conference, EWEC 2006, Athens, Greece, February-March 2006*.
 [4] Kariniotakis G, Halliday J, Brownsword R, et al. 2006 Next Generation Short-Term Forecasting of Wind Power—Overview of the ANEMOS Project. *In Proceedings of the European Wind Energy Conference, EWEC 2006, Athens, Greece, February-March 2006*.
 [5] Porter K, Rogers J 2010 Status of Centralized Wind Power Forecasting in North America; NREL/SR-550-47853; National Renewable Energy Laboratory: Golden, CO
 [6] Yang X, Xiao Y, Chen S 2005 Wind speed and generated power forecasting in wind farm *Proceedings of the CSEE* **25** 1-5

- [7] Ding M, Zhang L, Wu Y 2005 Wind speed forecast model for wind farms based on time series analysis *Electric Power Automation Equipment* **8** 32-4
- [8] Xiao Y, Wang W, Huo X 2007 Study on the time-series wind speed forecasting of the wind farm based on neural networks *Energy Conserv. Technol.* **2** 2
- [9] Wang H, Hu Z, Hu M, et al. 2012 Short-Term Prediction of Wind Farm Power Based on PSO-SVM *Power and Energy Engineering Conference (APPEEC) Asia-Pacific, 2012; IEEE: 2012* 1-4
- [10] Pousinho H, Catalão J, Mendes V 2010 Wind power short-term prediction by a hybrid PSO-ANFIS approach *MELECON 2010-2010 15th IEEE Mediterranean Electrotechnical Conference, 2010; IEEE: 2010* 955-60
- [11] Pratheepraj E, Abraham A, Deepa S N, et al. 2011 Very Short Term Wind Power Forecasting Using PSO-Neural Network Hybrid System *In Advances in Computing and Communications, Springer* 503-11
- [12] Eberhart R, Kennedy J 1995 A New Optimizer Using Particle Swarm Theory *In Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 4-6 October 1995* 39-43
- [13] Kennedy J, Eberhart R 1995 Particle Swarm Optimization. *In Proceedings of the IEEE International Conference on Neural Networks, 27 November-01 December 1995* 1942-8
- [14] Goldberg D E, Richardson J 1987 Genetic Algorithms with Sharing for Multimodal Function Optimization *In Proceedings of the Second International Conference on Genetic Algorithms and Their Application, Cambridge, MA, USA* 41-9
- [15] Brits R, Engelbrecht A P, Van den Bergh F 2002 A Nicheing Particle Swarm Optimizer *In Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning, Singapore* 692-6
- [16] Brits R, Engelbrecht A P, Van Den Bergh F 2003 Scalability of Niche PSO *In Proceedings of the 2003 IEEE Swarm Intelligence Symposium, April 2003* 228-34
- [17] Gao Y, Ren Z 2007 Adaptive Particle Swarm Optimization Algorithm with Genetic Mutation Operation *In Third International Conference on Natural Computation, Haikou, Japan, August 2007* 211-5
- [18] Faerman M, Birnbaum A, Berman F, Casanova H 2003 Resource allocation strategies for guided parameter space searches *Int. J. High Perform. C* **17** 383-402
- [19] Potter M A 1997 *The design and analysis of a computational model of cooperative coevolution* PhD. dissertation, George Mason University, Fairfax, VA, USA
- [20] McClelland J L, Rumelhart D E 1986 PDP Research Group *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, 2nd ed.; The MIT Press: Cambridge, MA, USA*
- [21] Hopfield J J, Tank D W 1986 Computing with neural circuits- a model *Science* **233** 625-33
- [22] Ogasawara E, Martinez L C, de Oliveira D, et al. 2010 Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series *In The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18-23 July 2010* 1-8
- [23] Sola J, Sevilla J 1997 Importance of input data normalization for the application of neural networks to complex industrial problems *IEEE Trans. Nuclear Sci.* **44** 1464-8
- [24] Louka P, Galanis G, Siebert N, et al. 2008 Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering *J. Wind Eng. Ind. Aerod.* **96** 2348-62
- [25] Libonati R, Trigo I, DaCamara C C 2008 Correction of 2 m-temperature forecasts using Kalman filtering technique *Atmos. Res.* **87** 183-97
- [26] Spiegel M 1998 *Schaum's Outline of Probability and Statistics*; McGraw-Hill: New York, NY, USA 157-77

Authors



Zhang Hong, born on February 16, 1971, Huai'an City, Jiangsu Province, China

Current position, grades: Fujian Normal University until 2009. Since 2009, has been studying a doctoral program at Southeast University.

University studies: electrical engineering from Huaiyin Institute of Technology in 1992. M.Eng in electrical engineering from Southwest Jiaotong University in 2006.

Scientific interest: power system computation and simulation.

Person-organization fit evaluation and process optimization based on the matching theory

Chengmeng Xue^{1, 2}, Yu Yang^{1*}, Tao Yang¹, Tingting Zeng¹

¹State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China

²School of Information, Guizhou University of Finance and Economics, Guizhou 550004, China

Received 12 May 2014, www.tsi.lv

Abstract

To achieve an optimal bidirectional person-organization fit (P-O fit) and improve the overall satisfaction degrees for both of the two sides, a bidirectional P-O fit evaluation and process optimization model is established based on the Matching Theory in this paper. To begin with, the bidirectional P-O fit evaluation factors set is built after the analysis of the indexes of these factors, and the index weights are calculated with the Rough Set Theory; Then, a Bidirectional P-O fit Evaluation and Process Optimization Model is proposed, with the Fit Conflict Resolve Algorithm (CRA) to ensure the persons and organizations to be matched one-to-one; Finally, the validity of this model is verified by its implementation in the enterprise HXMS.

Keywords: persons and organizations, bidirectional fit, fit degree, Cauchy Distribution Function

1 Introduction

The person-organization fit (P-O fit) issue has received more and more public attentions nowadays. The Fit Theory argues that the proper P-O fit can have a positive effect on the rapid and healthy development of enterprises, and it contributes to the synergistic interactions, reducing conflicts and optimizing the production efficiency as well. Currently, the P-O fit issue has become the basic principle of the organizational management with its vital role to improve organizational effectiveness and build harmonious organizations.

Organizations in enterprises indisputably rely on their human resources management (HRM) to take the most advantage of their human resources, and the corn part of HRM is the management for the persons and the organizations. That is to say, HRM aims essentially to work out a proper, if not optimal, P-O fit solution, which is the premise of the smooth operation of HRM [1]. The P-O fit actually embodies the consistency degree of P-O characteristics, leading how to evaluate and measure the P-O fit degree effectively and accurately to be a priority. Previous researches in this field can be classified into two directions. One is the opinion of Supplementary Fit and Complementary Fit proposed by Muchinsky and Monahan [2], Supplementary Fit here referring to the fit between person characteristics or attributes (individual values, goals, attitudes, abilities, etc.) and organizational characteristics or attributes (organizational culture, structure, values, goals, etc.); while Complementary Fit here meaning that persons and organizations are complementary in terms of their characteristics and supplies. However, the other direction, Needs-supplies

and Demands-abilities, is suggested by Cable and Judge [3]. Viewpoint of Needs-supplies points out if what the organizations supply can meet the needs, desires and preferences of their persons, the P-O fit can be achieved; while that of Demands-abilities regards if the persons have abilities demanded by the organizations, then the P-O fit can be achieved. Although these two P-O fit perspectives have already been researched in depth, both of them have failed to reflect the bidirectional P-O fit behaviour in an integrated way. On the contrary, they tend to define P-O fit from just one perspective and ignore the other one. Practically, various definitions are used to illustrate different fit operations so as to measure the P-O fit degree. Withal, Kristof [4] puts forward a luculent P-O fit conception model (Figure 1) based on previous studies. The P-O fit in this model is defined as the P-O consistency in case of any of the following conditions: (1) either the persons or the organizations, at least one of the two sides, can provide the resources required by their counterpart; (2) similar basic characteristics are shared by the two sides of persons and the organizations; (3) both of the two conditions above are available.

The P-O consistency, which is generally measured by the P-O fit degree, is used to study the P-O interaction relationship from the perspectives of persons and organizations [5-7]. The P-O fit evaluation is often regarded as a complex process with its multiple objectives, levels and uncertainties due to various factors from persons (their needs, attitudes, values, knowledge, skills, abilities, personality, etc.), organizations (their culture, rewards, motivation, innovation, etc.), and the environment (dynamics, complexity, uncertainty, etc.).

* *Corresponding author* e-mail: yuyang@cqu.edu.cn

For now, methods for P-O fit evaluation mainly focus on two forms: direct evaluation methods and indirect evaluation methods [8-10].

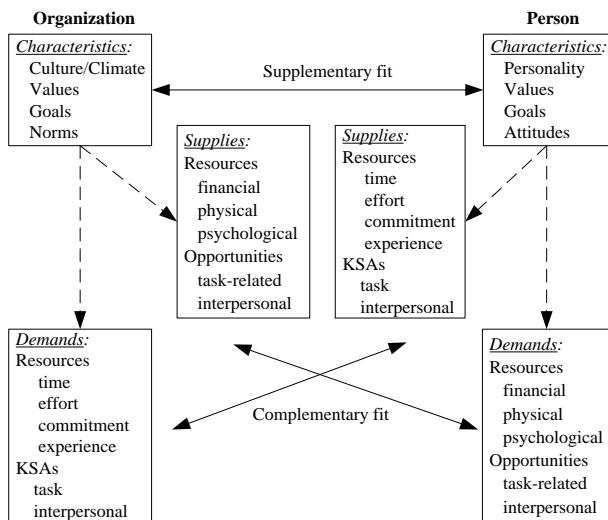


Figure 1 P-O fit model

Direct evaluation methods guide the persons to evaluate whether they have a proper P-O fit with their organizations, while indirect methods evaluate the persons' characteristics and organizations' characteristics under the same structure firstly and then compare the differences between the two sides with the method of D-score, Q methodology, Multi-regression methods and etc. Undesirable defects exist inevitably in these two methods thanks to their ignorance of the bidirectional P-O fit. The proper bidirectional P-O fit can on one hand stimulate the persons to improve their production efficiency with higher job satisfaction degrees, it can optimize the management level and organizational performance on the other hand, which make researches on approaches of bidirectional P-O fit necessary and beneficial. To resolve this issue, this paper focuses on the bidirectional P-O fit evaluation and process optimization at microscopic and macroscopic levels. First of all, the person evaluation factors set and the organization evaluation factors set are established respectively; Next, the P-O fit degree and the O-P fit degree are calculated respectively after persons and organizations make evaluations on their counterparts mutually; Then the integrated P-O fit degree is determined to measure the P-O consistency; Finally, the optimal bidirectional P-O fit result can be achieved to make sure the perfect fit of the organizations and the persons.

The rest parts in this paper are organized in the following way. Chapter 2 interprets the assumptions and notations presented in the research process. In chapter 3, two evaluation factor sets for persons and organizations are built separately to support the bidirectional P-O fit evaluation. Chapter 4 establishes the bidirectional P-O fit evaluation and process optimization model to calculate the integrated P-O fit degree and evaluate the bidirectional P-O fit process. This model is verified by an example in chapter 5. In the end, Chapter 6 makes a

conclusion of the contents, which are introduced in this paper.

2 Descriptions of the assumptions and notations

2.1 ASSUMPTION

This paper is compiled under the following assumptions:

1 In the P-O fit evaluation process, evaluations given by the organization managers and the persons are given objectively;

2 In the P-O fit evaluation process, choices of the organization managers and the persons are made rationally;

3 In the P-O fit evaluation process, the organizations and the persons are matched one-to-one.

2.2 INTERPRETATIONS OF THE NOTATIONS

- M_k : the person who fits the organization, $k=1,2,3,\dots, K$;
- F_r : the organization which fit the person, $r=1,2,3,\dots, R$;
- P_i : the i^{th} evaluation index of the person evaluation factors set, $i=1,2,3,\dots, m$;
- O_j : the j^{th} evaluation index of the organization evaluation factors set, $j=1,2,3,\dots, n$;
- W_i : the weight of the i^{th} evaluation index of the person evaluation factors set;
- W_j : the weight of the j^{th} evaluation index of the organization evaluation factors set;
- $S_{rk}^{(i)}$: the satisfaction degree from organization r to the person k ;
- $T_{kr}^{(j)}$: the satisfaction degree from person k to organization r ;
- S_{kr} : the P-O fit degree from person k to organization r ;
- T_{rk} : the P-O fit degree from organization r to person k ;
- ST_{kr} : the integrated P-O fit degree between person k and organization r .

3 P-O evaluation factors set and the calculations

3.1 P-O EVALUATION FACTORS SET

The bidirectional P-O Fit aims to study the causes and effects of the consistency between the persons and the organizations they belong to, which in fact reflect the relevant factors influencing the P-O fit process. The proper P-O fit reflects their mutual adaption and harmony, which not only have positive impacts on variables at persons' level, it also plays an important role in variables at organizations' level. Mutual impact relationship in the fit process can be classified into two categories: influences made by persons' characteristics on organizations and influences made by organizations on persons. The former one make differences mainly with their knowledge, skills and abilities, or whether their personality traits meet the needs of the organizations, whether their value views are consistent with that of the

organization, whether their job performances reach the organizations' relevant standards, whether their complex needs can be satisfied by the organization and so on, all of which affect the P-O consistency degrees and the result of the P-O fit. While the later influence mainly results from the following aspects: whether the resources the organizations provide meet the needs of their persons, whether their target is in line with that of their persons, whether their cultural are similar to that of their persons, whether their structure allows for the self-development of their persons and so on. Generally speaking, the P-O interaction reflects the principle of the mutual accommodation and development between organization and person. Therefore, this paper analyses the bidirectional P-O fit indexes and builds two evaluations factors sets (Table 1) to evaluate and optimize the P-O fit issue.

TABLE 1 P-O evaluations factors set

Person evaluation factors set		Organization evaluation factors set	
P	Evaluation index	O	Evaluation index
P ₁	Communication Skills	O ₁	Organizational Performance
P ₂	Personal Values	O ₂	Organizational Innovation
P ₃	Professional Knowledge	O ₃	Organizational Culture
P ₄	Decision-making Ability	O ₄	Organizational Strategy
P ₅	Computer Application Ability	O ₅	Organizations Working Environment
P ₆	Individual Personality	O ₆	Organizational Diversity
P ₇	Personal Attitudes	O ₇	Organizational Cohesion
P ₈	Continuous Learning Ability	O ₈	Organizational Norms
P ₉	Research Capacity	O ₉	Management Leading Level
P ₁₀	Adaptation Ability	O ₁₀	Employment Mechanism
P ₁₁	Innovation Ability	O ₁₁	Comprehensive Strength
P ₁₂	Sense of Responsibility	O ₁₂	Organizational Stability
P ₁₃	Healthy Status	O ₁₃	Organizational Image
P ₁₄	Leadership Skills	O ₁₄	Organization Working Content
P ₁₅	Interpersonal Skills	O ₁₅	Personnel Promotion Opportunities
P ₁₆	Education Background	O ₁₆	Personnel Learning Opportunities
P ₁₇	Personal Image	O ₁₇	Organizational Location

3.2 CALCULATION OF THE EVALUATION INDEX WEIGHTS

In the P-O fit evaluation process, the organizations and the persons vary in terms of their emphasis of the evaluation indexes, thus this paper calculates the evaluation index weights W_i and W_j by using the Rough Set Theory on the basis of the P-O evaluation factors set. The main contents are as follows:

For decision problems with more than one index, assume the condition attribute set is C and the decision attribute set is D. To facilitate the discussion, the relevant definitions are given firstly [11, 12]:

Definition 1 Assume the tetrad $I = (U, A, V, f)$ to be a decision information system. If $A = C \cup D, C \cap D = \emptyset$, then $I = (U, C, D, V, f)$ is the decision table, C is the condition attribute set and D is the decision attribute set.

Definition 2 In the decision information system $I = (U, A, V, f)$, $U = \{u_1, u_2, \dots, u_{|U|}\}$ is the Domain, standing for the samples data collection (object or entity) with $U \neq \emptyset$; A is a set of non-empty finite attributes, with $A = \{a_1, a_2, \dots, a_{|A|}\}$ presenting a collection of all attributes; $V = \bigcup_{a_j \in A} V_{a_j}$ with V_{a_j} presenting the attribute range of $a_j \in A$; $f: U \times A \rightarrow V$ is the information function, which is a single map to ensure the attribute of any object from U to have a unique information value, i.e., $\forall a_j \in A, \forall u_i \in U$ and $f(u_i, a_j) \in V_{a_j}$.

This paper exemplifies the weights calculation process of each index from the person evaluation factors set $P = \{P_1, P_2, \dots, P_{17}\}$ so as to calculate all the other index weights W_i with the Rough Set Theory, and the basic steps are as follows [13]:

1) Determine the evaluation index P_i and the integrated evaluation value D , and calculate the dependence degree $\gamma_{P_i}(D)$ through formula (1), the coefficient $\gamma_{P_i}(D)$ here representing the dependence degree between D and P_i .

$$\gamma_{P_i}(D) = \frac{\text{card}(POS_{P_i}(D))}{\text{card}(U)} \tag{1}$$

Here $\text{card}(\bullet)$ represents the radix of the set.

2) Calculate the weight of index P_i

The weight of index $P_i (P_i \in P)$ can be regarded as the decision-making changing degree if we remove index P_i from the evaluation system. The greater the change is the more significant index P_i is. Therefore, the importance value of P_i can be shown through formula (2)

$$\text{Sig}(P_i) = \gamma_{P_i}(D) - \gamma_{(P - \{P_i\})}(D) \tag{2}$$

3) Normalize the index weights

Normalize the importance value of each evaluation index, whose weight can be calculated through formula (3)

$$W_i = \frac{\text{Sig}(P_i)}{\sum_{i=1}^m \text{Sig}(P_i)} \quad i = 1, 2, 3 \dots m \tag{3}$$

Here, W_i is the weight of the i^{th} evaluation index.

Similarly, each evaluation weight W_j of the organization evaluation factors set can be calculated by following the steps above.

4 Bidirectional P-O fit evaluation and process optimization

4.1 BIDIRECTIONAL P-O FIT EVALUATION

Due to the prevalence of the discrepancy between the desired expectations the organizations have on their persons and the actual conditions the persons achieve, each organization has set a corresponding objective evaluation index system for their persons, known as "P-O fit degree". For this consideration, the satisfaction degrees can be divided into nine grades {extremely unsatisfied, fairly unsatisfied, unsatisfied, not very satisfied, basically satisfied, very satisfied, satisfied, fairly satisfied, extremely satisfied}, with its corresponding comment sets: { $u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9$ }. On this basis, the satisfaction evaluation matrix from person k to organization r and that from organization k to person r can be obtained separately (in Table 2 and Table 3), where $i = 1,2,3,\dots,m; j = 1,2,3,\dots,n; r = 1,2,3,\dots,R; k = 1,2,3,\dots,K$.

TABLE 2 Satisfaction evaluation from person k to organization r

	O	O ₁	O ₂	O ₃	O _j	O _n
Person k	Organization 1	u_5	u_3	u_7	...	u_1
	Organization 2	u_2	u_6	u_3	...	u_7
	Organization 3	u_1	u_3	u_9	...	u_8
	Organization r
	Organization R	u_4	u_1	u_2	...	u_3

TABLE 3 Satisfaction evaluation from organization r to person k

	P	P ₁	P ₂	P ₃	P _i	P _m
Organization r	Person 1	u_2	u_7	u_3	...	u_2
	Person 2	u_2	u_5	u_3	...	u_6
	Person 3	u_3	u_7	u_5	...	u_5
	Person k
	Person K	u_6	u_4	u_7	...	u_2

In reality, persons are more sensitive to the dissatisfaction degree than to the satisfaction degree. That is to say, various complaints can be made by the persons if the evaluations from organizations on their persons decrease a level; while a slight increase, if any, of the satisfaction degree can be obtained if the fit degree increase a level. Considering this, we take advantage of Cauchy Subjective Distribution Function [7-10].

$$f(u) = \begin{cases} [1 + \alpha(u - \beta)^{-2}]^{-1} & 1 \leq u \leq 5 \\ a \ln u + b & 5 \leq u \leq 9 \end{cases} \quad (4)$$

Here, α, β, a, b are undetermined coefficients. When the satisfaction degree is rated as "extremely satisfied", membership is 1 and $f(9) = 1$; when the satisfaction degree is rated as "basically satisfied", membership is 0.8 and $f(5) = 0.8$; when the satisfaction degree is rated as "extremely dissatisfied", the membership degree is 0.01 and $f(1) = 0.01$. α, β, a, b

can be calculated separately through equation (1), and then the language-based evaluation value from Table 2 and Table 3 can be transferred into a numeric type. Take the organization evaluation process for example; the satisfaction evaluation matrix after the transformation is expressed in Table 4.

TABLE 4 Satisfaction evaluation matrix post-transformed from person k to organization r

		O ₁	O ₂	O ₃	O _j	O _n	Integrated degree
	Weight	W ₁	W ₂	W ₃	W _j	W _n	
Person k	Organ. 1	$T_{k1}^{(1)}$	$T_{k1}^{(2)}$	$T_{k1}^{(3)}$	$T_{k1}^{(j)}$	$T_{k1}^{(17)}$	T_{1k}
	Organ. 2	$T_{k2}^{(1)}$	$T_{k2}^{(2)}$	$T_{k2}^{(3)}$	$T_{k2}^{(j)}$	$T_{k2}^{(17)}$	T_{2k}
	Organ. 3	$T_{k3}^{(1)}$	$T_{k3}^{(2)}$	$T_{k3}^{(3)}$	$T_{k3}^{(j)}$	$T_{k3}^{(17)}$	T_{3k}
	Organ. r	$T_{kr}^{(1)}$	$T_{kr}^{(2)}$	$T_{kr}^{(3)}$	$T_{kr}^{(j)}$	$T_{kr}^{(17)}$	T_{rk}
	Organ. R	$T_{kR}^{(1)}$	$T_{kR}^{(2)}$	$T_{kR}^{(3)}$	$T_{kR}^{(j)}$	$T_{kR}^{(17)}$	T_{Rk}

Here, T_{rk} presents the fit degree organization r has on person k , which can be calculated through formula (5)

$$T_{rk} = \sum_{j=1}^n T_{kr}^{(j)} W_j \quad j = 1,2,3, \dots, n \quad (5)$$

Similarly, the fit degree person k has on organization r can be calculated through formula (6)

$$S_{kr} = \sum_{i=1}^m S_{rk}^{(i)} W_i \quad i = 1,2,3, \dots, m \quad (6)$$

Therefore, the integrated fit degree ST_{kr} between organization r and person k can be calculated through formula (7)

$$ST_{kr} = \sqrt{S_{kr} T_{rk}} \quad r = 1,2,3, \dots, R; k = 1,2,3, \dots, K \quad (7)$$

In a summary, after sequencing the integrated fit degree ST_{kr} in a descending order, the fit result with the maximum ST_{kr} is the optimal P-O fit result for person K and organization r , meaning person k finds organization r the most satisfying and vice versa.

4.2 BIDIRECTIONAL P-O FIT CONFLICT RESOLVE ALGORITHM

According to assumption 3, the persons and the organizations should be matched one-to-one. However, it is often the case in reality that more than one person is selected by the same organization or an organization is selected by more than one person, which inevitably leads to bidirectional P-O fit conflicts. If not resolved properly, the fit conflicts are bound to destroy the organization's harmony and jeopardize the organizational efficiency.

To solve these fit conflicts, this paper comes up with a Fit Conflict Resolve Algorithm to achieve the process optimization.

- 1) List the desired persons of organization F_r based on their ST_{kr} values in a descending order, and the persons ranking the first few places are more desirable than their

followers by organization F_r . The sequence result is $M_1, M_2, M_3, \dots, M_k$;

2) Fit organization F_r to the person with the maximal ST_{kr} , where $r = 1, 2, 3, \dots, R$;

3) Detect the P-O fit conflicts. If none, then go to step (5); otherwise, go to step (4);

4) Organization F_r and F_v select the same person M_k (or person M_k and M_u select the same organization F_r). Compare the S_{kr} value of person K and organization F_r and the S_{kv} value of person K and organization F_v

- 1) If $S_{kr} > S_{kv}$, then M_k should select F_r ;
 - 2) If $S_{kr} < S_{kv}$, then M_k should select F_v ;
 - 3) If $S_{kr} = S_{kv}$, then M_k can select F_v or F_r at random.
5. Terminate the algorithm.

To sum up, the process of bidirectional P-O fit evaluation and optimization is shown in Figure 2.

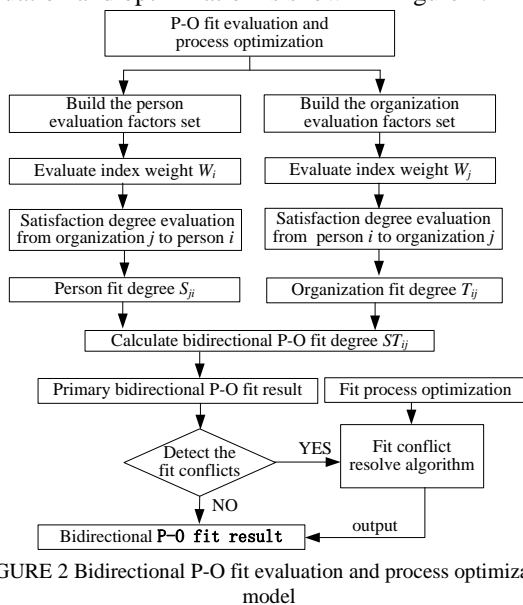


FIGURE 2 Bidirectional P-O fit evaluation and process optimization model

5 Case study

The method proposed in this paper is verified by an example in HXMS, a high-tech enterprise famous for its integration of the precision molds, precision stamping parts and plastic injection molding parts. With its ever-increasing development and rapid growth, HXMS plans to enlarge its recruitment scale in 2013 to deal with talent shortage issue in its expansion process. It launches activities to achieve a proper bidirectional P-O fit so as to get higher productivity efficiency and a better quality. Eight product development organizations and twelve persons are introduced in this paper as the objectives to verify the method proposed. The P-O fit evaluation index system is shown in Table 1. Specific analysis steps of bidirectional P-O fit evaluation and process optimization are as follows:

Step 1 Evaluate the index weights

Determine the index weights of P-O fit evaluation factors set. Calculation results are shown in Table 5 and Table 6.

TABLE 5 Index weights of the person evaluation factor set

Index	P ₁	P ₂	P ₃	P ₄	P ₅	...	P ₁₃	P ₁₄	P ₁₅	P ₁₆	P ₁₇
Weight W _i	.048	.061	.066	.070	.078070	.042	.064	.046	.040

TABLE 6 Index weights of the organization evaluation factors set

Index	O ₁	O ₂	O ₃	O ₄	O ₅	...	O ₁₃	O ₁₄	O ₁₅	O ₁₆	O ₁₇
Weight W _i	.042	.060	.069	.082	.073035	.047	.059	.054	.044

Step 2 Calculate the P-O fit degree and the O-P fit degree

1) Calculate the P-O fit degree

Firstly, construct the satisfaction evaluation matrix organization r for person K ; Next, construct the post-transferred satisfaction evaluation matrix through formula (4); Then the satisfaction degree $T_{rk}(k=1,2,3,\dots,12)$ from the eight development organizations on the twelve persons can be obtained through formula (5). The result is shown in Table 7.

TABLE 7 Sequence of T_{rk} (from organization r to person k)

No.	Tr1	Tr2	Tr3	...	Tr10	Tr11	Tr12
1	F ₃ 0.727	F ₂ 0.901	F ₈ 0.775	...	F ₁ 0.778	F ₇ 0.821	F ₈ 0.863
2	F ₅ 0.714	F ₅ 0.879	F ₃ 0.752	...	F ₅ 0.763	F ₂ 0.802	F ₆ 0.839
3	F ₇ 0.692	F ₆ 0.851	F ₄ 0.733	...	F ₄ 0.733	F ₅ 0.899	F ₂ 0.721
4	F ₄ 0.681	F ₇ 0.721	F ₂ 0.721	...	F ₂ 0.712	F ₁ 0.872	F ₇ 0.717
5	F ₁ 0.651	F ₄ 0.611	F ₅ 0.701	...	F ₆ 0.690	F ₄ 0.853	F ₅ 0.702
6	F ₈ 0.645	F ₁ 0.601	F ₁ 0.762	...	F ₇ 0.682	F ₃ 0.792	F ₁ 0.686
7	F ₆ 0.623	F ₈ 0.587	F ₆ 0.731	...	F ₃ 0.673	F ₆ 0.627	F ₃ 0.654
8	F ₂ 0.521	F ₃ 0.563	F ₇ 0.669	...	F ₈ 0.564	F ₈ 0.517	F ₄ 0.627

2) Calculate the O-P fit degree

Firstly, construct the satisfaction evaluation matrix person K on organization r ; Next, construct the post-transferred satisfaction evaluation matrix through formula (4). Then the satisfaction degree $S_{kr}(k=1,2,3,\dots,12)$ from the twelve persons to the eight organizations can be obtained through formula (6). The result is shown in Table 8.

TABLE 8 Sequence of S_{kr} (from person k to organization r)

No.	S _{k1}	S _{k2}	S _{k3}	...	S _{k6}	S _{k7}	S _{k8}
1	M ₃ .827	M ₂ .901	M ₈ .875	...	M ₁ .887	M ₇ .921	M ₈ .893
2	M ₅ .814	M ₅ .899	M ₁₀ .852	...	M ₉ .872	M ₁₀ .902	M ₆ .859
3	M ₇ .792	M ₆ .871	M ₄ .833	...	M ₃ .834	M ₅ .899	M ₁₂ .821
4	M ₄ .781	M ₇ .821	M ₂ .821	...	M ₅ .827	M ₁ .882	M ₇ .817
5	M ₁ .771	M ₄ .811	M ₁₁ .801	...	M ₁₀ .811	M ₄ .873	M ₅ .702
6	M ₁₀ .765	M ₉ .801	M ₁ .792	...	M ₇ .801	M ₉ .852	M ₁ .786
7	M ₆ .743	M ₁₁ .787	M ₆ .771	...	M ₂ .799	M ₁₂ .827	M ₉ .754
8	M ₂ .721	M ₃ .763	M ₉ .769	...	M ₄ .763	M ₃ .817	M ₁₁ .727
9	M ₁₁ .690	M ₁ .752	M ₁₂ .732	...	M ₈ .732	M ₆ .798	M ₁₀ .717
10	M ₉ .688	M ₈ .716	M ₃ .721	...	M ₆ .691	M ₈ .763	M ₂ .698
11	M ₈ .650	M ₁₂ .703	M ₃ .702	...	M ₁₁ .627	M ₂ .754	M ₃ .678
12	M ₁₂ .643	M ₁₀ .691	M ₇ .698	...	M ₁₂ .584	M ₁₁ .733	M ₄ .653

Step 3 Calculate the integrated P-O fit degree ST_{kr}
 Calculate the integrated P-O fit ST_{kr} through formula
 (7) based on the results from Table 7 and Table 8. Choose

organization r as the fit objective, the integrated P-O fit result between the twelve persons and the eight organizations are shown in Figure 3 (A-H).

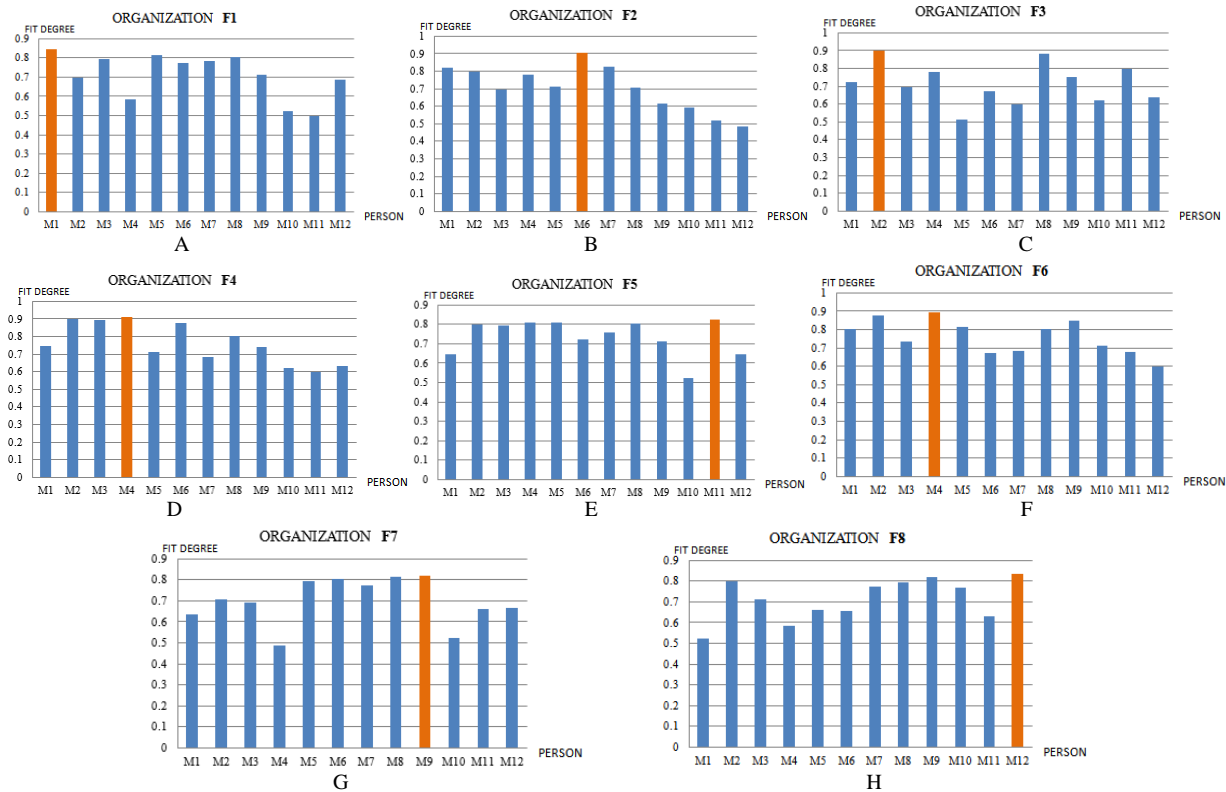


FIGURE 3 Integrated evaluation results between persons and organizations

Step 4 Resolve the P-O fit Conflicts

Figure 3 reveals a conflicting phenomenon where organization F4 and organization F6 regard persons M4 as their optimal objective simultaneously, a typical P-O fit conflict against assumption 3. Thus we resolve the fit conflict with the Fit Conflict Resolve Algorithm. Since $T_{44} > T_{64}$, person M4 should be the optimal for organization F4, and organization F6 should select person M2. Now that person M2 chooses organization F3, organization F6 is supposed to select person M5 finally.

Steps 5 Optimize the bidirectional P-O Fit result

According to the P-O fit evaluation and optimization process above, the final result of P-O fit for this enterprise is shown in Table 9.

TABLE 9 P-O bidirectional fit results

Organ. F_r	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
Person M_k	M_1	M_6	M_2	M_4	M_{11}	M_5	M_9	M_{12}
Fit degree	0.844	0.904	0.897	0.912	0.824	0.813	0.821	0.837

It can be seen from table 9 that the optimal P-O fit result for HXMS is M_1-F_1 , M_6-F_2 , M_2-F_3 , M_4-F_4 , $M_{11}-F_5$, M_5-F_6 , M_9-F_7 , $M_{12}-F_8$, which ensures the perfect bidirectional P-O fit.

6 Conclusions

The P-O fit issue, with its essential role in human resource management, exists widely in various management practices in almost all types of enterprises. The proper P-O fit have an overwhelming impact to make sure that the persons are attracted to their organizations with higher job satisfaction degrees and minimum demission rates. This paper studies the bidirectional P-O fit issues based on the Matching Theory, calculates the P-O fit degrees, the O-P fit degree and the integrated P-O fit by establishing the bidirectional P-O fit evaluation and process optimization model, thus to achieve a proper bidirectional P-O fit. Meanwhile, this paper provides references for the college students to make wise decisions during their job hunting process and guides the enterprises to seek for their perfect potential cooperative partners.





Acknowledgements

This research is supported by the National Natural Science Foundation, China (No.71071173, 71301176). Guizhou Science and Technology Department Research Project, China (Department of Guizhou Chimeric GY word [2012] 3050). Guizhou Province Soft Scientific Research Project, China (Department of Guizhou Chimeric R word [2013] LKC2030).

References

- [1] Tomoki Sekiguchi 2004 *Osaka Keidai Ronshu* 54(6) 179-96 (In Chinese)
- [2] Muchinsky P M, Monahan C J 1987 *Journal of Vocational Behavior* 31 268-77
- [3] Cable D M, Judge T A 1994 *Personnel Psychology* 47 317-31
- [4] Kristof A L 1996 *Personnel Psychology* 49 1-49
- [5] Cable D M, Judge T A 1997 *Journal of Applied Psychology* 82(4) 546-61
- [6] Becker B, Gethart B 1996 *Academy of Management Journal* 39(4) 779-801
- [7] O'Reil C A, Chatman, Caldwell D F 1991 *Academy of Management Journal* 34(3) 487-516
- [8] Verquer M L, Beehr T A, Wagner C H 2003 *Journal of Vocational Behaviour* 63 473-89
- [9] Kelly A, Piasentin Derek S C 2006 *Journal of Vocational Behaviour* 5 1-19
- [10] Karen J, Jansen A L 2006 *Journal of Managerial Issues* 18(2) 193-209
- [11] Yang J, Yang Y, Wang W L 2009 *China Mechanical Engineering* 20(20) 2452-05 (In Chinese)
- [12] Li Y Y, Yun J 2009 *Wuhan University of Technology (In Chinese)*
- [13] Yang Y, Liang X D, Li X L 2010 *Computer Integrated Manufacturing Systems* 16(5) 1020-06 (In Chinese)

Authors

	<p>Chengmeng Xue, born on November 9, 1965, Chongqing</p> <p>Current position, grades: Professor University studies: Chongqing University Scientific interest: Human Resource Management(HRM), product management, industrial engineering Publications: 3 Experience: C M Xue was born in Chongqing, China, in 1965. Currently, she is pursuing his Ph.D. degree in Mechanical Engineering at Chongqing University. His research interests include Human Resource Management(HRM), product management, industrial engineering, and collaborative product development.</p>
	<p>Yu Yang, born on September 10, 1971, Chongqing</p> <p>Current position, grades: Professor University studies: Chongqing University Scientific interest: customer collaborative product development, system reliability, product process management and so on. Publications: 30 Experience: Y Yang received his M.E. and Ph.D. degrees from Chongqing University, China, in 1995 and 1999, respectively. Now, he has been a professor in School of Mechanical Engineering, Chongqing University since 2003. During the year 2005-2006, he studies at Purdue University, USA. In 1997, he was selected for the New Century Excellent Talents in University program in China. He is currently working on customer collaborative product development, system reliability, product process management and so on. He is also the leader of National Natural Science Foundation of China (No. 71071173).</p>
	<p>Tao Yang, born on May 8, 1987, Anhui Province</p> <p>Current position, grades: Ph.D. students University studies: Chongqing University Scientific interest: product design, system reliability, Industrial Engineering(IE), and collaborative product development Publications: 3 Experience: Tao Yang was born in Anhui Province of China in 1985. He received the B.E. degree in Industrial Engineering from Shandong Institute of Business and Technology in 2010. Currently, he is pursuing his Ph.D. degree in Mechanical Engineering at Chongqing University. His research interests include product design and system reliability, industrial engineering, and collaborative product development.</p>
	<p>Tingting Zeng, born on July 23, 1987, Shandong Province</p> <p>Current position, grades: Postgraduate Student University studies: Chongqing University Scientific interest: Human Resource Management(HRM), Industrial Engineering(IE), product development Experience: Tingting Zeng Currently, he is pursuing his Master's degree in Mechanical Engineering at Chongqing University. Her research interests include product design and system reliability, industrial engineering, and collaborative product development.</p>

An artificial fish swarm algorithm for solving a bi-objective capacitated vehicle routing problem

Jinling Li^{1, 2}, Haixiang Guo^{1, 3*}, Yan Chen⁴, Deyun Wang¹, Kejun Zhu¹

¹School of Economics and management, China University of Geosciences, Wuhan, 430074, P.R. China

²Jiangcheng College, China University of Geosciences, Wuhan, 430200, P.R. China

³Key Laboratory of Tectonics and Petroleum Resources, China University of Geosciences, Ministry of Education, Wuhan Hubei, 430074, P.R. China

⁴School of Distance and Continuing Education, China University of Geosciences, Wuhan, 430074, P.R. China

Received 11 April 2014, www.tsi.lv

Abstract

The paper focuses on a capacitated vehicle routing problem with two objectives: one is attainment of specific load factor and the other is minimization of total travel cost. Our approach is based on artificial fish swarm algorithm, a swarm-based heuristic, which mimics the foraging behaviour of a fish swarm. After initializing a school of artificial fish, whose validity is guaranteed by a designed repair operator, global optimal solution search is processed through random behaviour, prey behaviour, swarm behaviour, and follow behaviour. Experimental results for a practical distribution instance are reported and show that the artificial fish swarm algorithm performs better than sweep algorithm and genetic algorithm. This paper contributes to the solution methods of vehicle routing problem.

Keywords: Vehicle routing problem, Artificial fish swarm algorithm, Sweep algorithm, Genetic algorithm

1 Introduction

In this paper, we present an artificial fish swarm algorithm for solving a bi-objective capacitated vehicle routing problem with attainment of specific load factor as the first objective and minimization of total travel cost as the second objective. The capacitated vehicle routing problem (CVRP) [1] is a static and basic version of the vehicle routing problem (VRP) [2]. Its objective is to find optimal routes for a fleet of m identical vehicles serving a set of n customers from a single depot. Each vehicle has a maximum capacity $Q_i (i=1, \dots, m)$. The demands q_j of the customers $j=1, 2, \dots, n$ are also deterministic and known in advance, and no split deliveries are available. A solution of the CVRP is described as a set of routes, each starting and ending at the depot and satisfying the conditions that each customer is visited only once and the accumulative demand of the customers in a same route for vehicle i limits to the capacity Q_i . A nonnegative cost c_{ij} originally based on the distance d_{ij} exists between a pair of customers (i, j) , contributing to the total travel cost which should be minimized. As an extension of the well-known traveling salesman problem (TSP), the CVRP is NP-hard so that only small-sized instances can be solved to optimality using exact algorithms [1, 3]. Thus, considerable problem-specific heuristics and meta-heuristic algorithms including the

sweep algorithm [4], the genetic algorithm [5-6], the tabu search [7], the artificial bee colony algorithm [8] and the ant colony algorithm [9] are introduced into the solution methods. However, to our knowledge, the artificial fish swarm algorithm [10] (AFSA) generally adopted for solving continuous problems is scarcely applied to CVRP. In this paper, we endeavor to expand the solution methods of CVRP by adopting AFSA, whose general procedure is illustrated in Figure 1. In real-life applications, the minimization of total distribution cost is often not the only objective. Various other aspects impact the quality of a solution [11]. Our work is just motivated by that kind of appeal from a company named Zhengzhou Coal and Electricity Materials Supply and Marketing Company (ZCEMS&M) in Henan province of China. The company devotes to convey dangerous goods from the depot to 14 coal mines. Without computation, the manager who is charging of distribution dangerous goods in ZCEMS&M used to assign the transport work arbitrarily and the decision making of a route is based on the manager's empirical knowledge. To cutback the enormous operation cost brought by transportation, the manager resorted to our team for a decision support system concerning vehicle routing. One of the constraints they proposed is that the vehicles launch to serve the set of customers only if the total demand exceeds a deterministic load, the percentage of the total demand to the capacity. Therefore, to generalize the problem, we formulate it as a bi-objective CVRP, namely the CVRP

* Corresponding author e-mail: faterdumk0732@sina.com

with minimum load constraint.

In our work, we use the artificial fish swarm algorithm [10] to solve the considered problem. Finally, we compare the result with that of another two algorithms (sweep algorithm and genetic algorithm) and verify the validity and robustness of AFSA to resolve CVRP.

The paper is organized as follows. Section 2 presents the mathematical model of the problem in this paper. In Section 3, different components of our algorithm is specified and integrated to operate. In Section 4, the algorithm is examined on a practical distribution instance, and the result is compared with that of the other two algorithms in the same circumstance. Section 5, eventually, gives concluding remarks and future research.

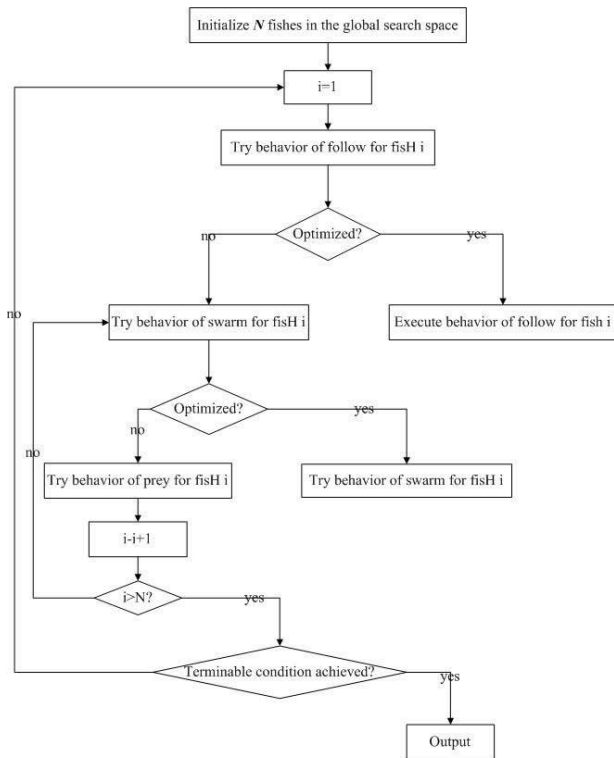


FIGURE 1 General procedure of AFSA

2 Model formulation

This section describes our model for the bi-objective CVRP. It can be seen as a metamorphosis of the model applied by Laporte et al. [12-13] and P. Reiter [14] for the DCVRP and CVRPB. The CVRP with two objectives in this paper (One is attainment of specific load factor and the other is minimization of total travel cost) is formulated as follows. The problem stretches itself on an undirected graph $G=(V,E)$, where $V=\{0,1,\dots,n\}$ is the set of vertices, namely the depot and customers, and $E=\{\{i,j\}:i,j\in V,i<j\}$ is the set of edges. Index 0 denotes the depot where m available vehicles $M=\{1,\dots,m\}$ of capacity $Q(v)$ (of which $v\in\{1,\dots,m\}$ and denotes a vehicle) are located. Meanwhile, the set of

customers is given as $V_0=V\setminus\{0\}$. Each customer i has a nonnegative demand $q(i)$. Moreover, to each edge $e\in E$, a cost value c_e , which can also be interpreted as the distribution time or as the length of edge e is associated. The cost matrix C is composed of all the cost values. We assume that C is symmetric and no service time are present. But the elements of C are not supposed to fulfil the triangle inequality (i.e., the distance function of every two customers is not a metric but given as a parameter).

For abbreviation, $\theta(S)$ signifies the set of edges in G with exactly one end vertex in S , e.g., $\theta(S)=\{\{i,j\}\in E:i\in S,j\in V\setminus S\}$, and $\theta(\{i\})$ is simply marked as $\theta(i)$. Furthermore, $\gamma(S)$ denotes the set of edges with both end vertices in S e.g., $\gamma(S)=\{\{i,j\}\in E:i,j\in S\}$. Finally, $(S:\bar{S})$ is the set of edges with one end in S and the other in \bar{S} . For each edge e , the decision variable x_e is defined as the multiplicity of edge e being assigned as part of a route, where $x_e\in\{0,1\}$ if e is not incident to depot, otherwise $x_e\in\{0,1,2\}$ (because the situation where some vehicle only serves customer i may exists, and thus edge $\{0,i\}$ occurs twice).

Additionally, ε denotes the specific load factor. For each vehicle v , the decision variable x_v is defined as whether the load of vehicle v exceeds load factor ε , of which $x_v=1$ if it is, otherwise $x_v=0$, and the decision variable x_r is defined as whether the load of vehicle r exceeds 0, of which $x_r=1$ if it is, otherwise $x_r=0$.

The considered bi-objective CVRP of this paper is provided by the following linear bi-objective optimization problem. The problem formulation contains the function $\pi(i,v)$, which is defined as whether vertex i is in the route of vehicle v , and equals 1 if it is, otherwise 0.

$$\min(\sum_{v\in M} Px_v + \sum_{e\in E} c_e x_e), \tag{1}$$

$$s.t. \sum_{e\in\theta(i)} x_e = 2 \quad \forall i\in V_0, \tag{2}$$

$$\sum_{e\in\theta(0)} x_e = 2 \sum_{r\in M} x_r, \tag{3}$$

$$\sum_{i\in V_0} q(i)\pi(i,v) \leq Q(v) \quad \forall v\in M, \tag{4}$$

$$x_v \in \{0,1\}, x_e \in \{0,1\} \quad \forall e \notin \theta(0)$$

$$x_e \in \{0,1,2\} \quad \forall e \in \theta(0)$$

Equation (1) defines the two objective functions to be minimized, of which P denotes a constant that is large enough. Equations (2) and (3) convince that exactly two edges are incident to each customer vertex and that accurately $2m$ edges are connected to the depot vertex. Equation (4), finally, restricts that the total demand of the customers that each vehicle serves cannot exceeds its capacity.

It should be noted that with constraint (3), the number of routes is a variable, i.e., when the total quantity of a set of demands is too small and a single vehicle can cover with a load approximating load factor ε , and other vehicles have no need to be launched and their loads equal 0. So, the vehicles available are scheduled dynamically.

3 Algorithmic solution

As a swarm-based heuristic, which mimics the foraging behaviour of a fish swarm, AFSA has outstanding performance in solving complicated practical problems, and becomes eye-catching for its simplicity of simulation. The algorithm searches for the global optimum value through individuals executing various behaviours.

In this section, firstly, separate components of the algorithm are explained, and then how they are woven together to solve the CVRP is described.

3.1 INITIAL SOLUTION

The artificial fish (AF) is the individuals that swim to find the optimal solution. Each AF holds a variable, namely the solution of the problem. Section 3.1.1 presents what is the solution looks like and how to formulate a solution. In Section 3.1.2, we proposed a repair operator to guarantee feasibility of solutions represented by AF.

3.1.1 Solution presentation

Each AF is denoted by a two-dimensional array $x[m][n]$, of which m is the quantity of vehicles and n is the quantity of customers. In this paper, we consider relatively large-scale vehicle routing problem, with more than 2 vehicles. The array can also shortly written as x . Figure 2 illustrates a representation of a CVRP instance with $m=4$ and $n=14$. For each value of the array, $x[i][j](i=1, \dots, m, j=1, \dots, n)$ denotes the j th customer that the vehicle i serves, e.g., $x[1][2]=4$ and $x[1][5]=0$, because vehicle 1 serves the fourth customer and the fifth customer does not exist.

2	4	7	9	0	0	0	0	0	0	0	0	0	0	0
8	6	3	11	1	0	0	0	0	0	0	0	0	0	0
14	5	12	13	10	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 2 Solution presentation

As Figure 2 shows, vehicle 1 serves customer 2, 4, 7, and 9, and its route can be denoted as $0-2-4-7-9-0$, of which 0 is the depot; vehicle 2 serves customer 8, 6, 3, 11, and 1, and its route can be denoted as $0-8-6-3-11-1-0$; vehicle 3 serves customer 14, 5, 12, 13 and 10, and its route can be denoted as $0-14-5-12-13-10-0$; vehicle 4 serves none of the customers and is not launched. Obviously, the solution satisfied the constraints of Equation (2) and Equation (3) in our model, see Section 2. To facilitate the following behaviours, we prescribe that all the customers should be located before the 0s, so the customer next to 0 is the last customer the vehicle visits.

To initialize a solution, we assign the $1 \sim n$ customers into a stochastic vehicle route. Therein, the customers are randomly ordered to assure the diversity of the fish swarm.

3.1.2 Repair operator

For AF x , when the total demand g of the customers in vehicle i exceeds the capacity $Q(i)$, we need to keep removing the customers backward from the last one until $g < Q(i)$, and record them into the single array $b[n]$. The locations that the removed customers used to sit are filled by 0s. As to the customers stored in array b , we reassign them to the other vehicles and the launched vehicles are at the first consideration.

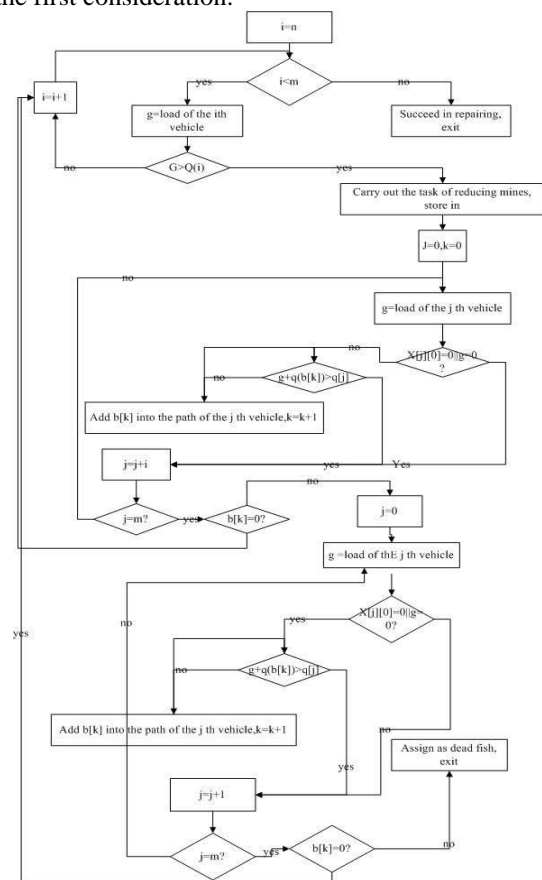


FIGURE 3 Procedure of repairing AF

Notes: $m = 4, n = 14, visual = n = 14,$

$$x = \begin{matrix} & 2 & 4 & 7 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & 6 & 3 & 11 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 14 & 5 & 12 & 13 & 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

However, it may happened that when all the vehicle are saturated, some customers are still not reassigned and are stored in array b . In this case, we deem AF x a dead fish. To isolate that dead fish, we put $1 \sim n$ customers into every row of the array x , so that it is permanently inferior to any live fish and cannot be perceived by the live ones. The concept of perception is explained in Section 3.3.1. Figure 3 illustrates the process of repair operator.

3.2 FOOD CONCENTRATION

We define food concentration as the objective functional values of AF. For abbreviation, $f(x)$ denotes the food concentration of AF x . Equation (5) represents the calculation process of $f(x)$, and the values of the decision variables on the right can be obtained according to the solution x . For detail of the right part, see the mathematical model in Section 2.

$$f(x) = \sum_{v \in M} P x_v + \sum_{e \in E} c_e x_e \tag{5}$$

3.3 BEHAVIOURS OF ARTIFICIAL FISH

The process of AF's optimization is achieved by three kinds of basic behaviours. This section discusses how to realize its perception and behaviours of prey, swarm, and follow.

3.3.1 Perception

Perception ensures that AF is able to perceive the state of other fishes and the surrounding environment. The scope of perception is limited, and $visual$ is the threshold. If the distance between AF x and xv , $d(x, xv)$, is bigger than $visual$, AF x cannot perceive xv . As to the distance d , we define it as the degree of similarity between two variables.

$$e_{ij} = \begin{cases} 1 & (xv_{ij} - x_{ij} \neq 0) \\ 0 & (xv_{ij} - x_{ij} = 0) \end{cases} \tag{6}$$

$$d(x, xv) = \sum_{i=0}^M \sum_{j=0}^N e_{ij} \tag{7}$$

Equation (6) and (7) denotes how to calculate the distance of AF xv to focal AF x . As we have designed them as arrays, each AF consists of $m \times n$ values located in their respective rows and columns. If the two values,

x_{ij} and xv_{ij} , both of which are in row i and column j , are not equal, e_{ij} equals 1 and the 1 adds to $d(x, xv)$, otherwise e_{ij} equals 0 and the 0 adds to $d(x, xv)$.

It can be seen from the definition of variable that each value of the variable is just a symbol, and does not share digital meaning, thus, the addition and subtraction of variables is meaningless to the distance calculation. Therefore, the general methods calculating the distance of multi-dimensional vector is not appropriate, and may distort the visual perception of AFs. However, the problem-specific method introduced in this paper latch on the essence of distance and is much more logical and the bigger the value of distance, the higher level of heterogeneity between the two variables. When the value is bigger than our given threshold $visual$, the two fish are invisible to each other. The worst condition of the distance calculation occurs when the two distribution schemes are totally dissimilar and the relative distance is $2n$ (The location of customers $1 \sim n$ are all different, then $d = n + n = 2n$). Therefore, $visual$ should be smaller than $2n$.

As values of a dead fish do not equal 0 in every location, its distance to any live fish should be no less than $m \times n - n$, and it is invisible for any live one. Meanwhile, behaviours of the live fish are perfectly isolated from the dead.

3.3.2. Behaviour of prey

We decompose the behaviour of prey into three procedures. Firstly, the focal AF x stochastically searches a point xv that is perceivable, legitimate, and feasible. As to guarantee perceivably, the number of customers that each vehicle serves and the route along which each vehicle travels of xv should be similar to that of x at a specific level so that $d(x, xv) \leq visual$. As Section 3.3.1 explained, during the comparison of each value in row i and column j of xv and x , a 1 is added to $d(x, xv)$ in one of the three conditions: (a) $x_{ij} \neq 0$ and $xv_{ij} = 0$; (b) $xv_{ij} = 0$ and $x_{ij} \neq 0$; (c) $xv_{ij} \neq 0$ and $x_{ij} \neq 0$, but $xv_{ij} \neq x_{ij}$. To simplify the illustration, we use s to signify the number of locations in condition (a), s' in condition (b), and \bar{t} in condition (c). In addition, t denotes the number of locations in the condition that $xv_{ij} \neq 0$, $x_{ij} \neq 0$, and $xv_{ij} = x_{ij}$. The distance then can be described by function $d(x, xv) = s + s' + \bar{t}$ and we can derive from the principle of the solution in Section 3.1.1 that $s + t + \bar{t} = n$ and $s' + t + \bar{t} = n$. Therefore, $s = s'$, $d(x, xv) = 2s + \bar{t} \leq visual$, and $0 \leq s \leq visual / 2$, i.e. there should be at least $\bar{s}_{min} = n - visual / 2$ locations where $xv_{ij} \neq 0$ and $x_{ij} \neq 0$. To formulate a xv , firstly we

should specify the locations in xv where we allocate the n customers by sequentially selecting \bar{s}_{\min} locations in xv where $x_{ij} \neq 0$ from the first column and randomly distributing the rest $(n - \bar{s}_{\min})$ locations in back of the selected locations. For instance, see Part I in Figure 4, where we present an instance with $visual$ equalling n and x equalling the solution presented in Figure 2.

Having fixed the locations, we can get the exact value of s , and thereby the interval of t should be $[\max(n + s - visual, 0), n - s]$. Secondly, we should specify the route of each vehicle, i.e. the values of each row in array xv , by randomly choosing $t_{\min} = \max(n + s - visual, 0)$ locations where $xv_{ij} \neq 0$ and $x_{ij} \neq 0$ and assigning xv_{ij} as x_{ij} , and sequentially inserting the rest $(n - t)$ customers that are stochastically ordered into the other blank locations. For instance, see Part II in Figure 4. For feasibility, the repair operator designed in Section 3.1.2 is introduced to remedy xv . Although $d(x, xv)$ might slightly exceed $visual$, we regard the excess acceptable to avoid the further sophisticated and nonessential repair operation.

Part I Location specifying

(1) Sequentially select $\bar{S}_{\min} = n - visual / 2 = 7$ location in xv

$$xv = \begin{matrix} - & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

(2) Randomly distribute the rest $(n - \bar{s}_{\min}) = 7$ locations in back of the selected locations

$$xv = \begin{matrix} - & - & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Part II Rout specifying

(1) Randomly choose $t_{\min} = \max(n + s - visual, 0) = 4$ locations where $xv_{ij} \neq 0$ and $x_{ij} \neq 0$ and assign xv_{ij} as x_{ij}

$$xv = \begin{matrix} - & 4 & - & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 6 & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 14 & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & - & - & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

(2) Arrange the rest $(n - t) = 10$ customers in stochastic order as

$$C_r = \{1, 2, 11, 3, 12, 5, 7, 8, 13, 10\}$$

(3) Sequentially inserts the customers in C_r into the blank locations

$$xv = \begin{matrix} 1 & 4 & 2 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 11 & 6 & 3 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 14 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 8 & 13 & 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

$$\therefore d(x, xv) = 12 < visual.$$

FIGURE 4 Instance of searching a perceivable point

Secondly, xv is estimated whether it is better than x . If it is, the values of variable xv will be recorded in another variable xn signifying the next status of x and originally equal the values of x . Otherwise, AF x repeats the first procedure until a better xv is obtained for try_number times at most. If after repeating try_number times a better xv still cannot be found, AF x will try the random behaviour, which means to exchange the locations of two customers, i and j , stochastically. Customer i and customer j can be in the same route of a vehicle or not, and the updated status will be recorded in xn . The random behavior prevents AFs from trapping into the local optimum and proceeds to the global optimum.

Thirdly, the food concentration of x after trying the behaviour of prey, namely $f(xn)$, is obtained. What to keep an eye on is that trying behaviours does not mean to update x with the optimized status but record the status in variable xn .

In the general artificial fish swarm algorithm applied to solve continuous problems, parameter $step$ is set to standardize the behaviours of AF. However, in this CVRP problem, every variable denotes a feasible distribution scheme, so the solution set is not continuous, but discrete, and the values of each variable are not numeric, but nominal. Obviously, the definition of $step$ is nonessential to this problem. Setting $step$, and enforcing AFs move in terms of $step$, would illegitimate the variables so that they need significantly complicated repair. Instead of step by step, we enable AFs to jump to better status directly.

3.3.3 Behaviour of swarm

The behaviour of swarm is condensed into four procedures. Firstly, the focal AF x finds n_f friends, all the other AFs perceivable, and records the variables into three-dimensional array $fri[fish_number][][]$, of which n_f denotes the quantity of AFs in the perception scope of x and $fish_number$ denotes the quantity of all AFs.

Secondly, we calculate the centre of these friends who is recorded as a temporary AF xc . For each location, i.e. row i and column j , we calculate the frequencies of 0

and each customer, ranging from 1 to n , and record the frequencies into single-dimensional array $\max[n+1]$. The value of $\max[r]$ signifies the frequency of customer r occurring in this location, and $\max[0]$ the frequency of 0. The customer or 0 who owns the biggest frequency is recorded in xc in the designated location, row i and column j . And the values of xc in other locations are calculated in the same way. Then, two illegal situations may happen to xc , where (a) some customers emerge repeatedly in various locations and (b) some customers never occur. Therefore, we propose a series of remedies.

To remedy situation (a), we record the frequencies of occurrence for $1 \sim n$ customers in locations of xc into single-dimensional array $b[n+1]$, of which the value of $b[i]$ denotes the frequency of customer i in xc , and $b[0]$ represents the frequency of 0; simultaneously, the locations in which a customer occurs again are assigned as 0, recording their row labels and column labels into the two-dimensional array $rc[n][2]$, of which $rc[][0]$ is the row label and $rc[][1]$ is the column label. After that, locations of xc with repeated customer are substituted for 0, and we can locate the places by reading data in array $rc[n][2]$.

To regulate situation (b), we scan values in array b and put customer i , of which $b[i]$ equals 0, in the locations stored in array rc . If some customers in the same boat remain, they are added to the launched vehicle arbitrarily. However, if the locations stored in array rc remain and happen to be in the middle of the route, xc violate the definition laws. A further remedy is necessary. Therefore, we move the customers after 0 forward, and the 0s before customer backward. Moreover, in order to cover the validity of xc , it is corrected by repair operator designed in Section 3.1.2.

Thirdly, we figure out whether (a) xc is superior to x and (b) it is not crowded around xc . If the condition above is satisfied, the values of xc are recorded in xn , otherwise xn remains the same to x . Condition (a) is based on comparing the food concentration of x and xc , namely $f(x)$ and $f(xc)$. Equation (8) denotes condition (b), i.e., if the food concentration of xc multiplied by the quantity of friends of x , n_f , is bigger than the saturation factor δ multiplied by the food concentration of x , AF x thinks that it is too crowded around xc . Equation (9) denotes that the saturation factor δ , one of the parameters of the algorithm, is determined by n_{\max} , the expectation of the maximum quantity of individuals we want to see near the extremum, and a , at what level we want the function value of individuals to approximate the extremum. For example, when we have determined that there should be at most 15 AFs around the target point whose food concentrations approximate that of the point at level of 80%, the saturation δ equals 12.

$$f(xc)n_f > \delta f(x), \tag{8}$$

$$\delta = an_{\max},$$

$$0 < a < 1. \tag{9}$$

Fourthly, the food concentration of xn , $f(xn)$, is returned, same to the last procedure of the behaviour of prey.

3.3.4 Behaviour of follow

As to the behaviour of follow, focal AF x firstly picks out the optimal one from its friends whose distance to x is within the perception scope, and we record values of the optimal friend and the quantity of the friends of x into a temporary AF xm and a numerical variable n_f respectively. Then, if xm is superior to x and it is not crowded around xm , namely $f(xm) < f(x)$ and $f(xm)n_f > \delta f(x)$, xm is duplicated by xn , otherwise xn remains the same to x . Finally, the behaviour of follow returns the food concentration of xn like the behaviours above.

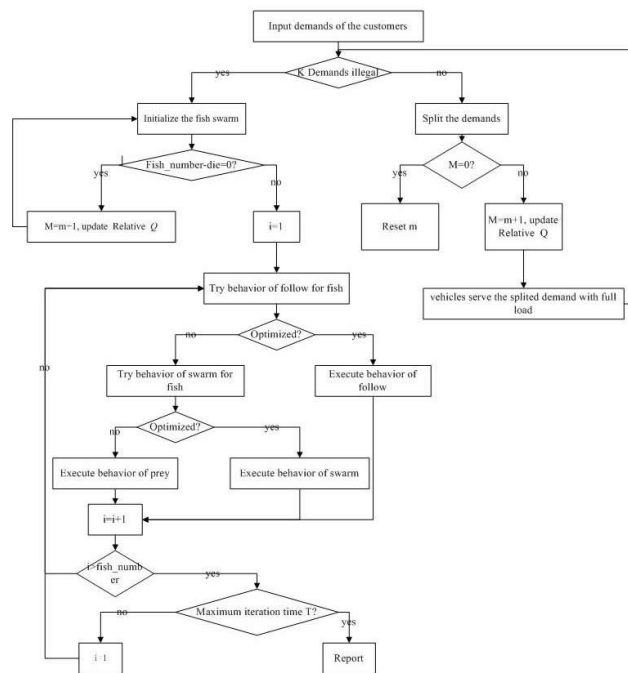


FIGURE 5 Complete flow of AFSA

3.4 EVALUATION OF THE BEHAVIORS

The evaluation of behaviours illustrates the logic of behaviour execution. For AF x , it is commended to try the behaviour of follow fist. If the food consistence returned by this behaviour is smaller, x exits the function

directly with the value of $f(xn)$ returned, otherwise tries the behaviour of swarm, and if the food consistence returned is smaller, x also exits the function directly with the value of $f(xn)$ returned, otherwise executes the behaviour of prey, and then exits with the value of $f(xn)$ returned. Variable xn is the next status of AF x , just as the sections of three behaviours clarified.

3.5 ACTION

Action means to update the focal AF x with xn . Although AF x tried many behaviours to optimize itself beforehand, its status has never been updated till now.

3.6 BULLETIN

Bulletin always holds the best record, variable $best_x$, throughout the algorithm. It is set to compare the variable x put into the bulletin function with $best_x$, based on their food concentrations. If x is superior to $best_x$, all the values of x will be duplicated to $best_x$ to guarantee the optimality of $best_x$.

3.7 INTEGRATION OF THE COMPONENTS

In this section, we illustrate how to conduct the entire process of the optimization. The algorithm this paper introduced progresses in four procedures. Firstly, we assign all the parameters, since they are the basic standard of the algorithm-processing environment and different sets of parameters lead to significantly distinct efficiency. The parameters are the number of vehicles and customers, m and n , the capacity of each vehicle, $Q(i)$, and the cost matrix, C . Figure 5 illustrates the entire flow of algorithm.

Procedure 1: Input of the demands. The deterministic set of demands activates the algorithm. However, in case the demands surpass the total load ability of vehicles, we should, firstly, judge the legality of the demands. In the algorithm, we tolerate a conditional extent of the demands exceeding the capacity. If a demand is bigger than the capacity, we split it by the capacity and cutback one of the vehicles, so the available vehicles is decreased to $m-1$. The surplus formulates a new demand used to update the demand of the original customer. With the decrease of available vehicles, m may equal 0, then a new circle of fleet scheduling begins and m is reset to the initial one. Then, go to Procedure 2.

Procedure 2: Initialization. For the best record $best_x$, we set it to the status of a dead fish. Then, the $fish_number$ fishes in the swarm are initialized, during which all the fishes are compared with $best_x$ in the process of Bulletin to update the best record. As to the dead fish formulated in period of initialization, we denote

the multiplicity of them as die . Therefore, the number of live fish is $fish_number - die$. If it equals 0, we have to add the vehicles that are available, otherwise go to Procedure 3. For that, we choose to circularly use the current vehicles, meaning the reused vehicle will serve twice.

Procedure 3: Optimization. The artificial fishes are in turns to action the behaviours based on the logic interpreted in evaluation of behaviours section, and update the bulletin instantly. While one cycle of that finishes, one turn of iteration accomplishes. When it adds to the maximum T times, the algorithm terminates.

Procedure 4: Reporting. Variable $best_x$ holds the optimal solution we find. To make it more comprehensive, we translate $best_x$ to distribution scheme that can be generally understood by the people outside this work.

4 Computational experiments

We have tested the AFSA algorithm on a practical instance of Zhengzhou Coal and Electricity Materials Supply and Marketing Company (ZCEMS&M). The experiment was performed on 1.86 Gigahertz computer, and the algorithm was coded in Visual C++ 6.0. The number of artificial fishes $fish_number$, the bound of perception $visual$, the times of independently searching try_number , the maximum iteration time T and the saturation factor δ are set as Table 1.

TABLE 1 Parameter Presentation

$fish_number$	$visual$	try_number	T	δ
50	16	20	5000	9

4.1 IMPLEMENTATION DETAILS

The dangerous goods distribution problem of ZCEMS&M can be described as that 4 vehicles that locate at a single depot serve 14 customers, namely the coal mines. That kind of vehicle routing problem takes place in the company every day, so the cost reduction brought by computerized optimization is fairly attractive to the company. The information of vehicles is presented in Table 2.

TABLE 2 Information of vehicles

The number of vehicles	4			
Capacity	2	2	2	2
Label	4545	4537	893	763
Oil consumption per kilometre	11.5	11.5	11	11

However, compared to the mathematical model, the oil consumption per kilometre of each vehicle $oil(i)(i=1, \dots, m)$ is added to the property of vehicle. We set oil price p as 6 yuan per litre. Relatively, the food

concentration calculation is changed by multiplying the total distance of each route of vehicle i by $6 \times oil(i)$. And the distances between each two customers are a general knowledge among the managers, shown in Table 3. The data serve to fulfil cost matrix C . The company claims that they are unwilling to launch a vehicle when the work load is under $5/6$ of the vehicle capacity.

To verify the validity of the algorithm, we test four sets of demands which are shown in Table 4. Then, one of the sets of demands is chosen to test by other algorithms, such as the sweep algorithm (SA) and the

genetic algorithm (GA). All of the results are shown in Section 4.2.

4.2 RESULTS ANALYSIS

Figure 6 shows the results of four sets of demands tested by AFSA. We can see that in each instance the optimal solution is obtained by iterating at most 260 times of set 2. The iterations of set 1, 3 and 4 are 22 times, 18 times, 66 times, respectively.

Distance (km)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	10	28	23	65	16	16	42	16	21	85	23	35	70	23
1	10	0	76	72	150	52	52	104	42	62	190	60	90	110	65
2	28	76	0	106	74	48	83	28	88	58	186	106	86	146	106
3	23	72	106	0	180	82	82	134	18	92	220	18	120	110	10
4	65	150	74	180	0	122	157	56	162	132	172	180	120	215	180
5	16	52	48	82	122	0	59	76	64	34	162	82	62	117	82
6	16	52	83	82	157	59	0	111	64	69	197	82	97	74	82
7	42	104	28	134	56	76	111	0	116	106	126	134	114	283	134
8	16	42	88	18	162	64	64	116	0	74	202	18	102	112	18
9	21	62	58	92	132	34	69	106	74	0	172	92	72	127	92
10	85	190	186	220	172	162	197	126	202	172	0	220	110	255	220
11	23	60	106	18	180	82	82	134	18	92	220	0	120	130	10
12	35	90	86	120	120	97	97	114	102	72	120	120	0	155	120
13	70	110	146	110	215	74	74	283	112	127	130	130	155	0	120
14	23	65	106	10	180	82	82	134	18	92	220	10	120	120	0

Notes: 1-14 denote the corresponding fourteen coal mines; 1 denotes Peigou; 2 denotes Daping; 3 denotes Zhanggou; 4 denotes Baiping; 5 denotes Micun; 6 denotes Chaohua; 7 denotes Gaocheng; 8 denotes Lugou; 9 denotes Laojuntang; 10 denotes Jinlong; 11 denotes Zhenxing; 12 denotes Cuimiao; 13 denotes Zhaojiazhai; 14 denotes Sanlimeiye; so do the following 1~14 in other figures and tables.

TABLE 4 Four sets of the demands

sets	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0.8	0	0.9	1.3	1.5	0	0	0.3	0.2	1	0	0.6	0.4
2	0.2	0.8	0.3	0.7	0.5	1	1	1	2	0.1	0.1	0.2	0.2	0.3
3	2	0.2	0.5	0.1	1.3	0.1	1.5	1.8	0	0	0.2	0.3	0.4	0.4
4	0.5	1	1.5	0	0.2	0	0.8	0.2	0	1	0	0.1	0.5	1.2

TABLE 5 Comparison results of the Algorithms

Algorithm	Solution	Total cost	Average load	launched vehicle
AFSA 1	0 -> 7 -> 2 -> 1 -> 0; 0 -> 8 -> 3 -> 13 -> 5 -> 0; 0 -> 10 -> 12 -> 14 -> 0. 0 -> 1 -> 8 -> 14 -> 0;	492.18	116.67%	3
AFSA 2	0 -> 3 -> 0; 0 -> 13 -> 10 -> 12 -> 0; 0 -> 7 -> 2 -> 5 -> 0. 0 -> 7 -> 2 -> 1 -> 0;	412.0	87.5%	4
SA	0 -> 3 -> 8 -> 0; 0 -> 14 -> 13 -> 5 -> 0; 0 -> 10 -> 12 -> 0.	452.55	87.5%	4
GA	0 -> 12 -> 5 -> 2 -> 7 -> 0; 0 -> 3 -> 0; 0 -> 14 -> 8 -> 1 -> 13 -> 0; 0 -> 10 -> 0.	465.87	83.75%	4

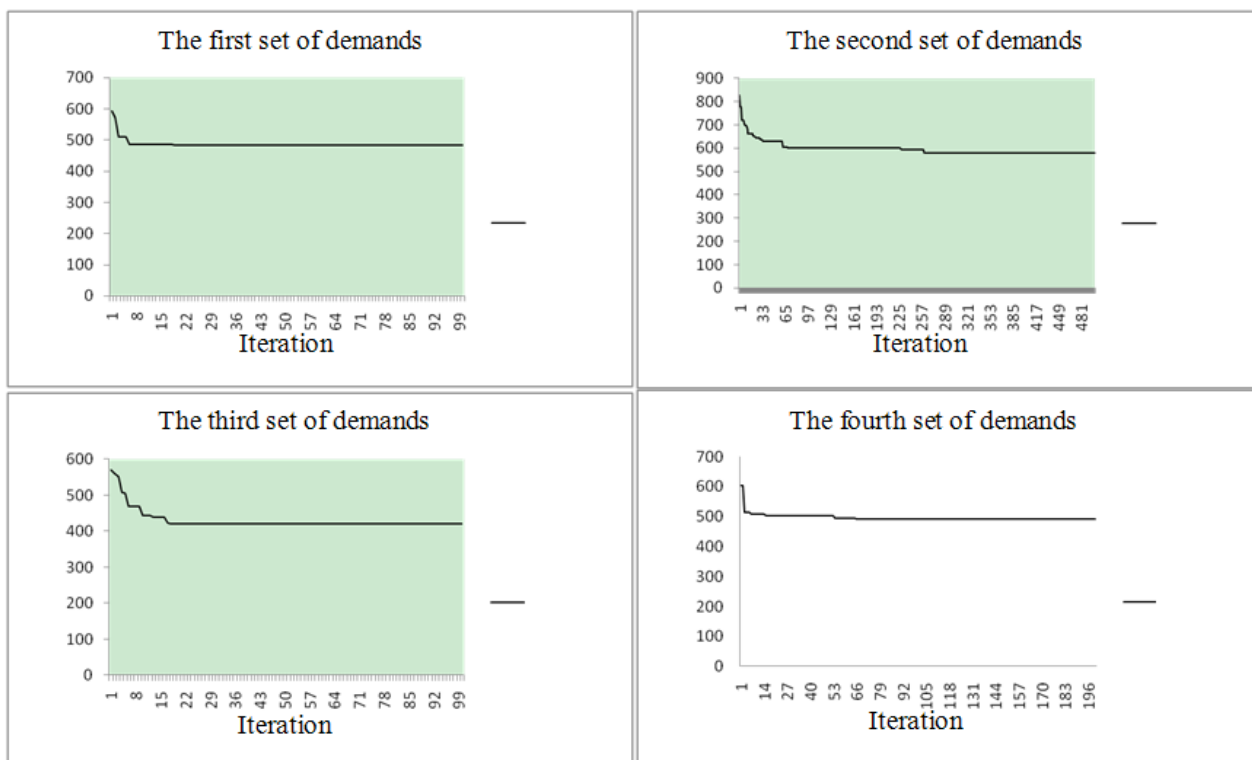


FIGURE 6 Four sets of demand test results of AFSA

The set of demands that we choose to test by other algorithms is the fourth, and Table 5 shows the results. Since the sweep algorithm and genetic algorithm have not consider the first objective of our model, to assure the comparability, we set the load factor of the algorithm as 0 to liberate the minimum load constrain and the artificial fish swarm algorithm in that situation is signed as AFSA 2, with AFSA 1 denotes the original standard. We can see from the table that only AFSA 1 completely satisfied our first objective, but the cost of AFSA 1 is higher than other algorithms. However, if we release the minimum load constrain, AFSA is significantly superior to other two algorithms. Furthermore, it implies that the company may suffer increased transportation cost by achieving the satisfied load.

5 Conclusions

In this paper, we present an artificial fish swarm algorithm (AFSA), a fairly new heuristic, for the capacitated vehicle routing problem (CVRP) with the minimum load constrain. The strategy and optimization process of the AFSA is not complicated and can be applied for practical problem solving appropriately. However, when coming across some general designs of components that violate the well-known laws of real world, the author is suggested to bravely abandon the trivial ones, or innovatively redesign them in terms of the specific problem. For example, we have discarded the



component *step* incident with behaviours of the artificial fishes in Section 3.3, and provided a new method of perception process and central point calculation in Section 3.3.1 and 3.3.3 for VRP solving by AFSA. This paper will continuously consider more actual restrictions such as the volume of goods, accidents during the distribution and emergency factors in order to enrich the content of VRP. Meanwhile, more intelligence algorithms can be applied as taboo algorithm, ant colony algorithm, artificial bee colony, etc.

Acknowledgements

This work is supported by the National Natural Science Foundation of China No.71103163, 71103164, 71301153, by Program for New Century Excellent Talents in University, No. NCET-13-1012, by Research Foundation of Humanities and Social Sciences of Ministry of Education of China No.10YJC790071, by the Fundamental Research Funds for National University, China University of Geosciences(Wuhan) No.CUG110411,CUG120111, G2012002A, CUG140604 by the open foundation for the research centre of resource environment economics in China University of Geosciences (Wuhan) and by the open foundation for Key Laboratory of Tectonics and Petroleum Resources (China University of Geosciences), Ministry of Education No. TPR-2011-11.

References

- [1] Toth P, Vigo D (Eds.) 2002 Models, relaxation and exact approaches for the capacitated vehicle routing problem *Discrete Applied Mathematics* **123** 487-512
- [2] Dantzig G B, Ramser J H 1959 The truck dispatching problem *Management Science* **10**(6) 80-91
- [3] Baldacci R, Toth P, Vigo D 2010 Exact algorithms for routing problems under vehicle capacity constraints *Annals of Operations Research* **175** 213-45
- [4] De Bacher B, Furnon V, Shaw P, et al. 2000 Solving vehicle routing problems using constraint programming and metaheuristics *Journal of Heuristics* **6** 501-23
- [5] Li Dawei, Wang Li, Wang Mengguang 1999 Genetic algorithm for vehicle routing problem with time windows *Systems Engineering – Theory & Practice* **8** 65-9 (in Chinese)
- [6] Brysy O, Dullaert W 2002 A fast evolutionary metaheuristic for the vehicle routing problem with time windows *International Journal of Artificial Intelligence Tools* **12**(2) 143-57
- [7] Ma Huawei, Yang Shanlin 2008 Improved tabu search algorithm for vehicle routing problem with alternative time windows *Journal of System Simulation* **20**(16) 4454-7
- [8] Szeto W Y, Yongzhong Wu, Sin C Ho 2011 An artificial bee colony algorithm for the capacitated vehicle routing problem *European Journal of Operational Research* **215** 126-35
- [9] Bell J E, McMullen P R 2004 Ant colony optimization techniques for the vehicle routing problem *Advanced Engineering Informatics* **18** 41-8
- [10] Li Xiaolei, Shao Zhijiang, Qian Jixin 2002 An optimizing method based on autoumous animats: fish-swarm algorithm *Systems Engineering – Theory & Practice* **22**(11) 32-6 (In Chinese)
- [11] Jozefowicz N, Semt F, El-Ghazali Talbi 2008 Multi-objective vehicle routing problems *European Journal of Operational Research* **189** 293-309
- [12] Laporte G, Desrochers M, Nobert Y 1984 Two exact algorithms for the distance constrained vehicle routing problem *Networks* **14** 161-172
- [13] Laporte G, Mercure H, Nobert Y 1986 An exact algorithm for the asymmetrical capacitated vehicle routing problem *Networks* **16** 33-46
- [14] Reiter P, Gutjahr W J 2012 Exact hybrid algorithms for solving a bi-objective vehicle routing problem *Central European Journal of Operations Research* **20** 19-43

Authors	
	<p>Jinling Li, born in April, 1981, Wuhan City, Hu bei Province, P.R. China</p> <p>Current position, grades: Lecture of Jiangcheng College, China University of Geosciences. University studies: Management Science and Engineering. Scientific interest: Study on mathematical model, data mining. Publications: more than 6 papers published in various journals.</p>
	<p>Haixiang Guo, born in September, 1978, Wuhan City, Hu bei Province, P.R. China</p> <p>Current position, grades: Professor of China University of Geosciences. University studies: Management Science and Engineering. Scientific interest: Soft computing. Publications: more than 30 papers published in various journals.</p>
	<p>Yan Chen, born in September, 1976, Wuhan City, Hu bei Province, P.R. China</p> <p>Current position, grades: the lecturer in China University of Geosciences. University studies: Knowledge-based Systems and Group decision and Negotiation. Scientific interest: Decision Making, Comprehensive evaluation and Uncertainty. Publications: more than 5 papers published in various journals.</p>
	<p>Deyun Wang, born in September, 1983, Wuhan City, Hu bei Province, P.R. China</p> <p>Current position, grades: Associate professor in China University of Geosciences. University studies: Information system management. Scientific interest: Schedul optimization. Publications: more than 10 papers published in various journals.</p>
	<p>Kejun Zhu, born in October, 1953, Wuhan City, Hu bei Province, P.R. China</p> <p>Current position, grades: Professor of China University of Geosciences. University studies: Mathematics Scientific interest: Decision Making, Comprehensive evaluation and Uncertainty Publications: more than 100 papers published in various journals.</p>

Analysis of price rate models for household water consumption in urban China

Lijuan Wang¹, Song Jin², Tianwei Zhang^{1*}, Peter Chung³

¹School of Traffic and Transportation Shijiazhuang Tiedao University, Shijiazhuang City, Hebei Province, China, 050043

²Department of Foreign Language Shijiazhuang Tiedao University, Shijiazhuang City, Hebei Province, China, 050043

³LQCCD Research Center of Environmental Simulation and Pollution Control, Tainan, Taiwan, 72557

Received 26 March 2014, www.tsi.lv

Abstract

Price rate models are used for a variety of purposes including water conservation, what's more, it is an effective and feasible method to establish a water-conserving society. The survey of household water consumption was conducted in a residential community in Hebei Province. An analysis of the survey data shows that an average family's water consumption is 6.4t per month, the standard is level 3. Three price rate models are proposed, the principle of the first model is the decrease of water consumption in level 1 and no increase in level 2, the second model is no change in level 1 and an increase in level 2, the third model is no increase for 80% of families and only 20% increased. After an analysis and comparison of water consumption and expenses corresponding to the three models, this paper presents these three models as a valuable reference for department decisions. The results of the first model should be adopted in Hebei Province, the demarcation point of every level is 8t and 13t, the price of every level is 3.04 RMB/t, 4.56 RMB/t and 9.06 RMB/t.

Keywords: Price rate model, Resident water consumption, Data survey, Water conservation

1 Introduction

China is lacking in water resources, the average possession of water resource per person is only 1/4 of the worldwide average, especially in urban areas. In order to alleviate this supply and demand contradiction, many scholars have suggested the use of economic leverage to manage water consumption [1]. Amidst the many suggestions, the price rate model has been widely accepted. Historically, determination of water price in China has been under government control, low prices of water for urban water supply enterprises has resulted in years of losses [2]. The government supports the normal operation of the urban water supply but this also carries with it a heavy financial burden.

The "Document about Promoting Water Price Reform and Water Conservation" was issued by the State Council on April 19th, 2004. The document states that price rate models should be implemented no later than the end of 2005 in China. However, the implementation of this document has not been as ideal as supposed even today.

The National Development and Reform Commission (NDRC) and the Ministry of Housing and Urban-Rural Development (MHUD) have issued "Introductory Advice about Establishing Stepped Water Pricing for Urban Chinese Households" (hereinafter referred to as "Introductory Advice") on January 3rd, 2014. The "Introductory Advice" states that a reasonable price rate

system should be established, it's precondition is to ensure the availability of needed water.

The "Introductory Advice" proposes that all related regions should implement a system of price rate models by the end of 2015. The water consumption amount should be divided into at least three levels. The first water consumption level should cover 80% of urban residents' monthly water usage, and the second level should cover 95% of urban resident's monthly water usage. The price rate ratios for the three levels are to be set at no less than the ratio of 1:1.5:3. The suggested amount of level 1 water consumption is 2.6 t per person per month in Hebei province. The suggested amount of level 2 water consumption is 4.3 t per person per month.

2 Analysis of the price rate of water consumption

The price rate of water consumption is determined by the number of levels in the current rate system and the price of each level. Research suggests that a price rate system for water conservation significantly improves the efficiency of water use and allocation of resources [3, 4]. The relationship between water consumption and water price is shown in Figure 1. The number of levels is set at 3.

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

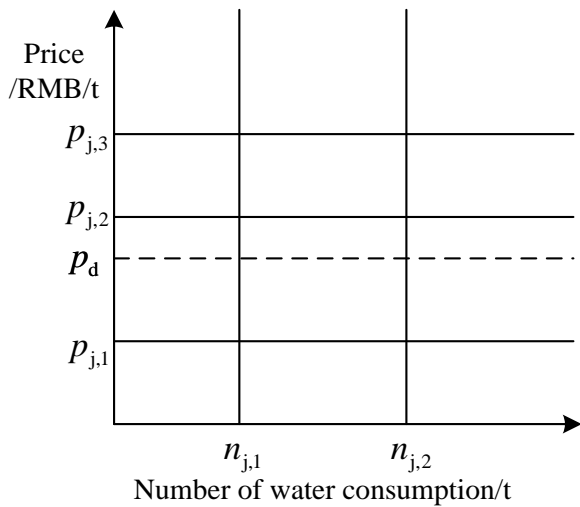


FIGURE 1 The relationship between water consumption and water price

In Figure 1, p_d is the current price, $p_{j,1}$, $p_{j,2}$ and $p_{j,3}$ are the prices of the first, second and third levels respectively. $n_{j,1}$ and $n_{j,2}$ are the upper limits of the first and second water consumption levels respectively. Therefore, the price rate of water consumption is determined by the values of five variables when the p_d is given.

According to “Introductory Advice”, the price of each level is estimated by the following equations:

$$p_{j,2} \geq 1.5p_{j,1} \tag{1}$$

$$p_{j,3} \geq 2p_{j,2} \tag{2}$$

$$p_{j,3} \geq 3p_{j,1} \tag{3}$$

However, if the only given value is p_d , the other five variables can not be obtained. In order to obtain these values, some other parameters should be acquired by the survey of urban residents’ water consumption [5].

3 Water consumption survey of urban residents in Hebei province

Water consumption of a residential community in Hebei Province was surveyed in December 2013. The survey sample was 815. Table 1 shows the ratio of residents’ water consumption.

According to Table 1, 80% of all household’s water consumption is less than 8t, and 95% of all household’s water consumption is less than 10t. The data substantiates that water consumption and its proportions satisfies normal distribution, the average value of water consumption is 6.4 t per month. The plotted result is shown in Figure 2.

TABLE 1 The ratio of residents’ water consumption

Water Consumption	Households	Ratio/%	Cumulative Ratio/%
(0,2]	15	1.84	1.84
(2,3]	28	3.44	5.28
(3,4]	48	5.89	11.17
(4,5]	123	15.09	26.26
(5,6]	158	19.39	45.64
(6,7]	149	18.28	63.93
(7,8]	125	15.34	79.26
(8,9]	97	11.90	91.17
(9,10]	38	4.66	95.83
(10,15]	24	2.94	98.77
(15,20]	10	1.23	100.00
(20,+∞)	0	0.00	100.00

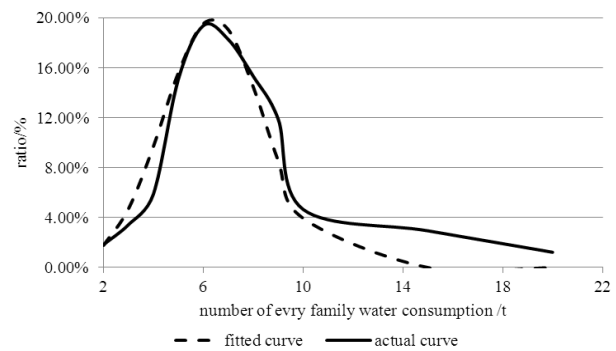


FIGURE 2 The plot of water consumption and proportion

4 Methodological approach and model construction

Based on the survey, there are very few households with a water consumption beyond 10 t per month, therefore, the number of levels beyond 3 is unnecessary. Three levels are suitable.

According to “Introductory Advice”, the upper limit of first-level water consumption is 8t, which is in agreement with actual survey results. What’s more, it is almost the same as advised values 7.8t. The upper limit of second-level water consumption is 13t. The water consumption of 95% of households is less than 10t, and it is also less than 12.9 t which is given in “Introductory Advice”. Because of the fluctuation of water consumption and the limited samples, the value can be selected as 13 t. The third-level water consumption is beyond 13 t. Models M1 and M2 are developed under these conditions.

Another method is used to confirm the upper limit of first-level water consumption. The average water consumption is 6.4 t, the value can be set at 7 because of rounding. The upper limit of second-level water consumption is 13t. With these parameters, the model M3 is established.

When the price rate is used in water consumption, the expenses should decrease for those whose water consumption is less than the upper limit of the first-level [6-7]. The expenses should not increase for those whose water consumption is between the upper limit of the first-level and the second-level. The model M1 is given by:

M1: s.t. $p_{j,1} < p_d$, (4) $p_{j,3} \geq 2p_{j,2}$, (13)

$p_{j,2} > p_d$, (5) $p_{j,3} \geq 3p_{j,1}$, (14)

$p_{j,1} \times n_{j,1} + p_{j,2} \times (n_{j,2} - n_{j,1}) = p_d \times n_{j,2}$, (6)

$p_{j,2} \geq 1.5p_{j,1}$, (7)

$p_{j,3} \geq 2p_{j,2}$, (8)

$p_{j,3} \geq 3p_{j,1}$, (9)

where Equation (4) expresses the first-level price rate is less than the current price, and Equation (5) and Equation (6) express the second-level price rate is higher than the current price. Equation (7-9) are the same as Equation (1-3).

Substituting $p_d = 3.63$, $n_{j,1} = 8$ and $n_{j,2} = 13$ in M1, the following results can be obtained: $p_{j,1} = 3.04$, $p_{j,2} = 4.56$, $p_{j,3} = 9.06$. Both Equation (10) and Figure 3 show the result.

$$P_1 = \begin{cases} 3.04n_1 & n_1 \leq 8 \\ 24.32 + 4.56(n_1 - 8) & 8 \leq n_1 \leq 13 \\ 47.12 + 9.06(n_1 - 13) & n_1 \geq 13 \end{cases} \quad (10)$$

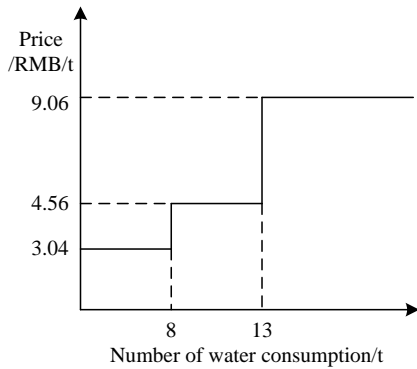


FIGURE 3 The result of model M1

Model M2 is based on the following hypothesis: the expenses of those whose water consumption is less than the upper limit of the first-level is equal to the current price, and the expenses of those whose water consumption is between the upper limit of the first-level and the second-level should increase. The model M2 is given by:

M2: s.t. $p_{j,1} = p_d$, (11)

$p_{j,2} \geq 1.5p_{j,1}$, (12)

where Equation (11) expresses the first-level water price is equal to the current price. Equation (12-14) are the same as Equation (1-3).

Substituting $p_d = 3.63$, $n_{j,1} = 8$ and $n_{j,2} = 13$ in M2, the following results can be obtained: $p_{j,1} = 3.63$, $p_{j,2} = 5.45$ and $p_{j,3} = 10.89$. Both Equation (15) and Figure 4 show the result.

$$P_2 = \begin{cases} 3.63n_2 & n_2 \leq 8 \\ 29.04 + 5.45(n_2 - 8) & 8 \leq n_2 \leq 13 \\ 56.29 + 10.89(n_2 - 13) & n_2 \geq 13 \end{cases} \quad (15)$$

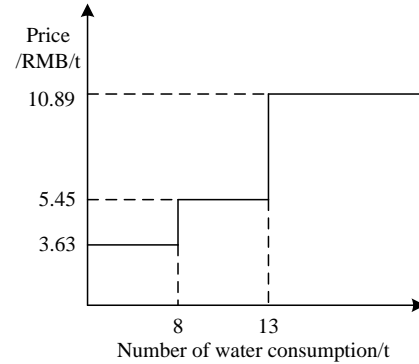


FIGURE 4 The result of model M2

The expenses of those whose water consumption is below the average are decreased, and the expenses of 80% of all households are not increased, while the expenses of 20% of all households are increased. The M3 is given by:

M3: s.t. $p_{j,1} < p_d$, (16)

$p_{j,2} > p_d$, (17)

$p_{j,1} \times n_{j,1} + p_{j,2} \times (n_{80\%} - n_{j,1}) = p_d \times n_{80\%}$, (18)

$p_{j,2} \geq 1.5p_{j,1}$, (19)

$p_{j,3} \geq 2p_{j,2}$, (20)

$p_{j,3} \geq 3p_{j,1}$, (21)

where Equation (16) expresses the first-level water price is less than the current price, and Equation (18) and Equation (19) express that the second-level water price is higher than the current price. Equation (19-21) are the same as Equation (1-3).

Substituting $p_d = 3.63$, $n_{j,1} = 7$, $n_{80\%} = 8$ and $n_{j,2} = 13$ in M3, the following results can be obtained: $p_{j,1} = 3.42$, $p_{j,2} = 5.13$ and $p_{j,3} = 10.26$. Both Equation (22) and Figure 5 show the result.

$$P_3 = \begin{cases} 3.42n_3 & n_3 \leq 7 \\ 23.94 + 5.13(n_3 - 7) & 7 \leq n_3 \leq 13 \\ 54.72 + 10.26(n_3 - 13) & n_3 \geq 13 \end{cases} \quad (22)$$

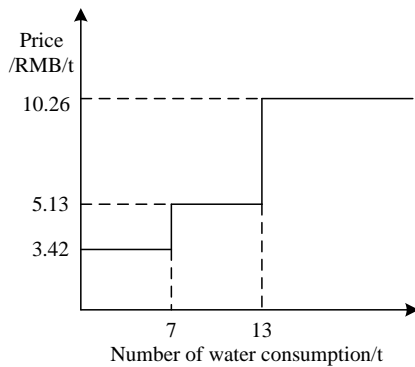


FIGURE 5 The result of model M3

5 Result Analysis

This paper establishes three models (M1, M2 and M3) based on different goals. Figure 6 shows each result, comparing it with the current price.

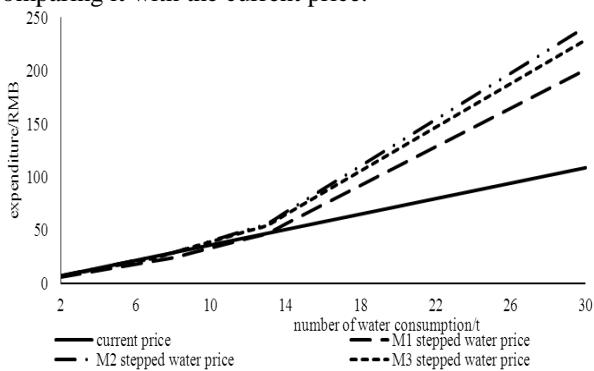


FIGURE 6 The comparison of models M1 M2 and M3

The following values shown in Table 2 are given to help more clearly comprehend the results.

According to Figure 6 and Table 2, water expenses are not increased when water consumption is below the first-level consumption. What's more, water expenses are decreased when water consumption is less than the first-level upper limit, even if the maximum increase of 25.41 RMB only accounts for incoming 0.93%.

When water consumption reaches the second-level, water expenses are slightly increased, even if the stepped water price of M2 is only increased by 9.1 RMB. The maximum increase of 56.29 RMB only accounts for incoming 2.1%.

TABLE 2 Water expenses at different water price rates

Water Consumption/t	Current Price/RMB	Water Price of M1/RMB	Water Price of M2/RMB	Water Price of M3/RMB
2	7.26	6.08	7.26	6.84
3	10.89	9.12	10.89	10.26
4	14.52	12.16	14.52	13.68
5	18.15	15.20	18.15	17.10
6	21.78	18.24	21.78	20.52
7	25.41	21.28	25.41	23.94
8	29.04	24.32	29.04	29.07
9	32.67	28.88	34.49	34.20
10	36.30	33.44	39.94	39.33
11	39.93	38.00	45.39	44.46
12	43.56	42.56	50.84	49.59
13	47.19	47.12	56.29	54.72
14	50.82	56.18	67.18	64.98
15	54.45	65.24	78.07	75.24
16	58.08	74.30	88.96	85.50
17	61.71	83.36	99.85	95.76
18	65.34	92.42	110.74	106.02
19	68.97	101.48	121.63	116.28
20	72.60	110.54	132.52	126.54
21	76.23	119.60	143.41	136.80
22	79.86	128.66	154.3	147.06
23	83.49	137.72	165.19	157.32
24	87.12	146.78	176.08	167.58
25	90.75	155.84	186.97	177.84
26	94.38	164.90	197.86	188.10
27	98.01	173.96	208.75	198.36
28	101.64	183.02	219.64	208.62
29	105.27	192.08	230.53	218.88
30	108.90	201.14	241.42	229.14

When water consumption reaches the third-level, the water expenses are increased quickly. Taking 30 t as an example, this paper presents three results. When water expenses were close to or over 200 RMB, compared with the monthly income of a family, which was more than 8000 RMB, the proportion of them are higher than 2.5%. However, compared with the monthly income of a family of 5000 RMB, when the water consumption reaches 20 t, the proportion of water expenses is higher than 2%. Xu notes that the residents started to care about water rate, attention to water conservation when the ratio between water expenses and family income is more than 2% [8]. Considering the public welfare of water resources and the goals of stepped water pricing, this paper presents that the result of model M1 is more suitable.

6 Conclusions and recommendations

Stepped water pricing is a widely accepted method for solving the problem of water waste. The NDRC and the MHUD have both issued documents stating that all related regions should carry out stepped water pricing before the end of 2015. With this in mind, this paper analyses the stepped water and pricing problem of Hebei Province, while at the same time establishing three models. The result of model M1 is regarded as the most suitable. However, other models can also be used as a reference for future studies.

Further research needs to be conducted to deal with more factors, and to seek out other possible ways. The best method and analysis remains relatively difficult, and so improvements in this area are also welcomed.

References

- [1] Zhao L G, Lin J, Wang X Y 2013 Studies on potential water resources crisis based on STIRPAT model: a case from Zhejiang in China *Nature Environment and Pollution Technology* **12**(4) 631-6
- [2] Chen H, Yang Z F 2006 Scalar urban water pricing model based on utility function *Resources Science* **28**(1) 109-112
- [3] Wang L F, Chen C X, Xing T 2011 An econometric model of multistep water pricing for urban household water consumption and its application *Journal of Yangtze River Scientific Research Institute* **28**(5) 5-8
- [4] Zhang D Z, Chen X Q 2003 Thinking on determination of domestic water price in Chinese cities *Journal of East China Normal University* (2) 81-5
- [5] Hari P S, Sharma M R, Quamural H, Naved A 2010 Sustainability of traditional drinking water sources in Himachal Pradesh *Nature Environment and Pollution Technology* **9**(3) 587-92
- [6] Chen Y D 2001 Sustainable development and management of water resources for urban water supply in Hong Kong *Water International* **26**(1) 119-28
- [7] Patamanska G, Chehlarova S S 2012 Changing existing irrigation systems and management in Bulgaria for sustainable use of water *INMATEH - Agricultural Engineering* **38**(3) 73-8
- [8] Xu Z C 2002 The reform project of the price of water supply in Shijiazhuang City *The Economy of Water* (5) 23-6

Authors



Lijuan Wang, born in May, 1980, Hebei Province, P.R. China

Current position, grades: Lecturer of School of Traffic and Transportation, Shijiazhuang Tiedao University, China.
University studies: B.Sc. and M.Sc. in Transportation Engineering from Shijiazhuang Tiedao University in China.
Scientific interest: engineering economy, transportation planning.
Publications: more than 10 papers published in various journals.
Experience: teaching experience of 10 years, 2 scientific research projects.



Song Jin, born in November, 1972, Hebei Province, P.R. China

Current position, grades: Lecturer of Department of Foreign Language, Shijiazhuang Tiedao University, China.
University studies: B.A. in ESL Education from Hebei Normal University and International Trade in University of International Business and Economics in China. M.A. from Hebei Normal University in China.
Scientific interest: economy, translationology.
Publications: more than 10 papers published in various journals.
Experience: teaching experience of 18 years, 5 scientific research projects.



Tianwei Zhang, born in December, 1979, Henan province, P.R.China

Current position, grades: Lecture of School of Traffic and Transportation, Shijiazhuang Tiedao University, China.
University studies: B.Sc. in Transportation Engineering from Lanzhou Jiaotong University in China. M.Sc. from Beijing Jiaotong University in China.
Scientific interest: optimization model, management of railway transportation.
Publications: more than 20 papers published in journals such as Journal of China Railway Society, Journal of Traffic and Transportation Engineering.
Experience: teaching experience of 12 years, 6 scientific research projects.

Calculation of China's environmental efficiency based on the SBM model with undesirable outputs

HaiLei Zou*, Cheng Wang

School of Science, China JiLiang University, HangZhou City, ZheJiang Province, China, 310018

Received 9 January 2014, www.tsi.lv

Abstract

With the rapid development of china's economy, the environment of china faces some prominent questions, industrial pollution, water pollution, serious smog and other problems continue get worse, therefore, it's very significant to analyse China's regional environmental efficiency. This paper mainly measure china's regional environmental efficiency by SBM model with undesirable outputs, and according to the calculation results, we find that the overall average level of china's environment efficiency is low, and the gap between different provinces is large. Finally, this article gives some policy proposal about how to increase china's environmental efficiency and reduce pollution emission.

Keywords: Environmental Efficiency, SBM Model, Undesirable Outputs

1 Introduction

Some countries have taken a lot of ways to solve environmental problems, but Global warming, water pollution and other problems are becoming worse, Environmental problems have become very serious in China, this not only affects people's health, but also seriously affected the china's sustainable development plan. On 12 August 2009, the state council approved a plan entitled "Environmental Impact Assessment Ordinance", This evaluation method has a high value in strategic planning. However China's environmental assessment is only just beginning, There are many deficiencies in the practical application, Therefore, establishing a set of scientific management evaluation method, improving the utilization of resources, reducing the emissions of waste materials will be the theme of environmental science at present, Since China is a major carbon dioxide and other greenhouse gas emission producer, China's environmental policy greatly impacts the balance of the global climate. In order to achieve the goal of energy conservation and emissions reduction, the provinces will require coordination, on the other hand, to determine the environmental efficiency of different regions accurately requires scientific and systematic environmental efficiency evaluation research. So that it can reflect environmental performance differences and gaps in different regions of China, each area can improve the environmental performance by using an objective reference standard. In the past few decades, many scholars studied the econometric analysis model and method to solve complex environmental problems. Among them, Environmental efficiency of computing problem has become a very important topic for many

scholars. However, how to find a more effective method to provide some quantitative information of performance evaluation and policy analysis remains an interesting area that should be thoroughly studied.

The structure of this article is as follows. Section 2 is a literature review. Section 3 introduces environmental efficiency evaluation model, data and indexes. Section 4 is the conclusions from the environmental efficiency values of china's provinces. And some relevant Suggestions are given in Section 5.

2 Literature review

DEA was proposed by Charns [1] and is a non-parametric statistical method based on linear programming model to measure the efficiencies of decision-making units. Subsequently, Fare developed the method and combined it with environmental production technology [2]. DEA based on the concept of "Relative Efficiency Evaluation", is the systematic analysis method using relative theory and a model of operational research to evaluate the relative efficiency of decision making units according to the multiple index input and output data. In recent years, DEA has gained great popularity in measuring the energy and environment efficiency at the macro-economy level. Zhou et al. summarized more than 100 DEA applications in energy and environment policy and they pointed out DEA model is characterized by its reference technology and efficiency measure [3]. Because China has already become the world's second largest economy and the largest Co emitter, it has become the focus with researchers studying the energy issue from the macro-economy perspective. Examples of such studies include [4–9], and so on. In the empirical part, DEA and

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

the corresponding extended model have been widely used in the energy and environmental efficiency evaluation. In the micro family part, Barba-Gutiérrez [10] analyses the environmental efficiency of household appliances and electronic equipment [10]. In the industrial part, Azadeh et al. [11], separately using the DEA model and the comprehensive model based on DEA, have explored both the environmental and energy efficiency of Iran's power transmission system and more than 100 countries' power production systems [11]. Mukherjee [12] has built a non-radial efficiency calculation method based on the DEA model, and Riccardi et al. [13], on the basis of the non-radial efficiency model, take undesirable outputs into consideration in the energy efficiency calculation. Mukherjee [12] and Riccardi et al. [13] separately study the energy efficiency of India's manufacturing industry and the global cement industry. In the macroscopic part, Chiu and Wu [14] have analysed the environmental efficiency of 27 Chinese provinces (municipality or autonomous regions) with the DEA model. With the enhancement of environmental protection consciousness [15], some environmental issues such as air pollution and hazardous waste have been widely recognized as a social issue. So, how to correctly measure environmental efficiency considering undesirable output becomes a hot topic.

3 Methodology and models construction

By constructing DEA models decision makers cannot only optimally design their systems, but also determine their optimal budgets. Deviation may appear in measurement of efficiency if CCR model is directly used and slack effect of input factors is neglected. Hence, Tone suggested a SBM model considering slack measurement, which is neither radial nor oriented, making it effective in remedying disadvantages of CCR and BCC models.

Traditional CCR model is expressed as follows:

$$\begin{aligned} &\min \theta, \\ &\text{s.t. } \sum_{j=1}^n \alpha_j x_{ij} + s^- = \theta x_0, i = 1, 2, \dots, m \\ &\sum_{j=1}^n \alpha_j y_{rj} - s^+ = y_0, r = 1, 2, \dots, s \\ &\alpha_j, s^-, s^+ \geq 0, j = 1, 2, \dots, n \end{aligned}$$

where x_{ij} is the inputs, y_{ij} is the outputs, x_0 and y_0 represent the input and output of DMU_0 respectively. s^- is the slacks of inputs; s^+ is the slacks of outputs. When the optimal value $\theta^0 = 1$, DMU_0 can be called as a weakly efficient DMU. When the optimal value $\theta^0 = 1$ and $s^- = s^+ = 0$, DMU_0 can be called as an efficient

DMU. If a new constraint, $\sum \alpha = 1$, is added in the model, we can gain the BCC model.

We further suppose that there are n DMUs with one input vector X and two output vectors for each DMU; desired output is Y^g and undesirable output is Y^b . The three vectors are defined as:

$$\begin{aligned} X &= (x_1, x_2, \dots, x_n) \in R^{m \times n}, x_i \in R^m, \\ Y^g &= (x_1^g, x_2^g, \dots, x_n^g) \in R^{s_1 \times n}, y_i^g \in R^{s_1}, \\ Y^b &= (x_1^b, x_2^b, \dots, x_n^b) \in R^{s_2 \times n}, y_i^b \in R^{s_2}. \end{aligned}$$

Here, $x_i > 0, y_i^g > 0, y_i^b > 0$. With constant return to scale, possible production volume is defined as P :

$$P = \{(x_i, y_i^g, y_i^b) / x_i \geq X\lambda, y_i^g \leq Y^g\lambda, y_i^b \geq Y^b\lambda, \lambda \geq 0\}.$$

So, SBM model can be expressed as:

$$\min \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s^-}{x_{i0}}}{1 + \frac{1}{s_1 + s_2} \left(\sum_{r=1}^{s_1} \frac{s_r^g}{y_{r0}^g} + \sum_{r=1}^{s_2} \frac{s_r^b}{y_{r0}^b} \right)}, \tag{1}$$

$$\text{s.t. } x_0 = X\lambda + s^-,$$

$$y_0^g = Y^g\lambda - s^g,$$

$$y_0^b = Y^b\lambda + s^b,$$

$$\lambda \geq 0, s^- \geq 0, s^g \geq 0, s^b \geq 0.$$

Thereinto, λ is weighing vector; s^- is slack variable of input, representing surplus of input; s^g, s^b are slack variables of output, representing inadequacy of output and surplus of undesirable output. Target function ρ' strictly descends in relation to s^-, s^g, s^b and $0 \leq \rho' \leq 1$. For a specific DMU, if $\rho' < 1$, the DMU is inefficient; if $\rho' = 1$ and $s^- = s^g = s^b$, the DMU is efficient. Unlike traditional CCR and BCC models, SBM model directly adds slack variables into target function, which cannot only solve the problem of slack in input and output but measure efficiency with undesirable output. As slack in input and output is considered in SBM model, Tone concluded that only when a DUM is efficient based on CCR model, it will be efficient based on SMB model with efficiency value in CCR bigger than or equal with that in SBM. In addition, SBM model is neither radial nor oriented, making it possible to avoid deviation caused by radial nor orient. Accordingly, nature of efficiency measurement can be better presented with this model.

This paper mainly investigates environmental efficiency conditions of the whole of China and each province under the pollution situation. Therefore, by using the related literature, we take the panel data of

mainland China’s 29 provinces from 2000 to 2012 as a sample.

For the input indicators, modern western economics considers capital, labour and land as the most important production input elements. According to its general meaning in economics and the majority of research, capital stock at constant prices is considered to be a good proxy indicator for capital investment. As Chinese official statistic bureaus have not publicly announced data of capital stock, it was estimated by using perpetual inventory method and expressed as follows:

$$K_t = K_{t-1}(1 - \delta_t) + I_t,$$

there into, K_t denotes capital stock in the year t ; I_t denotes investment in the year t ; δ_t denotes depreciation rate of fixed assets in that year.

It is generally thought that the earlier the benchmark is, the smaller the error will be in the long-term estimation. Therefore, in this article, we began our estimation from the year 1958. First, the sum of fixed assets in each year for each province was converted into an invariable value in 1958 by using an implicit deflator index. Then capital stock was calculated by using a basic equation of perpetual inventory method according to the benchmark of capital stock and depreciation rate set in advance. The value of capital stock for each province in our article is calculated on the basis of an invariable value in 1958.

$$\text{Capital stock in 1958} = \frac{\text{Sum of stock in 1959}}{\text{Depreciation rate} + \text{AIRIFA}},$$

There into, AIRIFA means average increasing rate of investment in fixed assets during 1958–1962; depreciation rate is universally 10.12% for all provinces and relevant data are from Statistic Yearbook of China for each year.

The number of employed persons at each year’s end in different provinces instead of the labour force indicators is selected; because each province’s land is nearly constant and cannot reflect their differences, we choose energy consumption, which has a closer connection with the production process instead of the land indicators. Generally speaking, the production process is the course to cost resources for outputs. The

criterion to evaluate DMU’s efficiency is fewer inputs and more outputs, which provides the DMU’s greater effectiveness. However, the production process products not only include the expected outputs, but also outputs such as pollutants, which people do not expect. Therefore, we divided outputs as desirable and undesirable. The GDP of different provinces will be used to present the desirable output. There is less detailed data of pollutants such as CO_2 , SO_2 and NO_2 in China. The emission of three wastes (industrial wastewater, gas and solid) in each province will be used as the similar proxy variables on behalf of the undesirable output. The data comes from the China Statistical Yearbook, China Energy Statistical Yearbook, and China Environment Statistical Yearbook, published by the National Bureau of Statistics of China. The basic statistical characteristics of specific input–output data are shown in Table 1.

4 China’s environmental efficiency

Environmental efficiency of each province from 2001 to 2012 is calculated by using Equation (1) and the results are shown in Table 2.

In line with convention, China is divided economically into three main regions: the East, the Central, the West. The East region includes: Beijing, Fujian, Guangdong, Hebei, Jiangsu, Shandong, Shanghai, Tianjin and Zhejiang, Heilongjiang, Jilin and Liaoning; the Central includes: Anhui, Henan, Hubei, Hunan, Jiangxi and Shanxi; the West includes: Gansu, Guangxi, Guizhou, Inner Mongolia, Ningxia, Qinghai, Shanxi, Chongqing, Sichuan, Xinjiang and Yunnan.

Environmental efficiency of the three regions is different greatly. The average value of the East is about 0.634, the Central is about 0.245 and a litter better of 0.323 for the West. The east has advanced technology and higher management level, And investment in terms of environmental management is quite huge, always controlling pollution at very low levels. Though the West drops behind economically, Western industry is at a stage of development, leading to relatively low emissions, in the west of the environment is better than the central, because the central of the low level of economic and the environmental pollution is serious.

TABLE 1 Descriptive statistical characteristics of input and output variables

variable	Input index			Output index	
	Capital x_1	Labour x_2	Energy x_3	GDP y^a	Waste y^b
Mean	2357.14	2265.56	6613.43	5043.89	7128.68
Median	1323.46	1879.67	5234.76	2879.76	4758.65
Maximum	20293.33	6348.32	34357.87	43879.45	56456.34
Minimum	27.18	210.00	176.98	67.46	112.87
Std.dev.	1467.94	1345.65	4367.76	3047.86	5423.65
Skewness	2.5472	0.6566	1.9235	2.6776	2.45
Kurtosis	11.7648	2.7453	7.5465	14.3424	12.5436

Note: Each sample has 600 observations, for the panel data includes 5 indicators of 29 provinces from 2001 to 2012

TABLE 2 Values of environmental efficiency in China (2001–2012)

year	2001	2002	2003	3004	2005	2006	2007	2008	2009	2010	2011	2012
Beijin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hebei	0.486	0.464	0.442	0.428	0.405	0.394	0.394	0.364	0.351	0.342	0.332	0.319
Shanxi	0.241	0.250	0.262	0.242	0.224	0.220	0.230	0.223	0.222	0.221	0.270	0.228
Innermongolia	0.357	0.336	0.335	0.325	0.331	0.268	0.243	0.217	0.204	0.198	0.221	0.222
Liaoning	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Jilin	0.476	0.454	0.472	0.494	0.471	0.442	0.415	0.398	0.364	0.323	0.313	0.287
Heilongjiang	0.336	0.446	0.382	0.391	0.375	0.392	0.381	0.405	0.413	0.391	0.384	0.342
Shanghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Jiangsu	0.558	0.535	0.524	0.498	0.483	0.454	0.412	0.375	0.364	0.362	0.364	0.360
Zhejiang	0.735	0.692	0.653	0.612	0.604	0.562	0.508	0.512	0.504	0.409	0.472	0.468
Anhui	0.296	0.289	0.298	0.286	0.262	0.277	0.274	0.262	0.264	0.247	0.257	0.245
Fujian	1.000	1.000	1.000	1.000	1.000	1.000	0.586	0.667	0.715	0.689	0.654	0.653
Jiangxi	0.241	0.233	0.214	0.212	0.197	0.183	0.175	0.162	0.158	0.152	0.152	0.149
Shandong	0.540	0.524	0.463	0.451	0.423	0.412	0.381	0.354	0.343	0.342	0.318	0.317
Henan	0.321	0.324	0.317	0.310	0.297	0.282	0.263	0.251	0.237	0.211	0.208	0.198
Hubei	0.362	0.368	0.367	0.363	0.365	0.346	0.311	0.333	0.329	0.321	0.319	0.318
Hunan	0.322	0.384	0.416	0.418	0.393	0.332	0.289	0.324	0.341	0.312	0.313	0.298
Guangdong	0.456	0.418	0.401	0.386	0.364	0.354	0.346	0.341	0.342	0.347	0.329	0.318
Guangxi	0.376	0.378	0.378	0.346	0.338	0.312	0.246	0.298	0.307	0.284	0.303	0.286
Hainan	0.758	0.748	0.752	0.682	0.683	0.645	0.665	0.752	0.723	0.741	0.698	0.701
Sichuan	0.158	0.157	0.164	0.165	0.560	0.155	0.156	0.144	0.195	0.144	0.195	0.185
Chongqing	0.156	0.163	0.156	0.163	0.423	0.153	0.152	0.142	0.185	0.142	0.186	0.187
Guizhou	0.128	0.131	0.134	0.128	0.125	0.114	0.112	0.113	0.124	0.116	0.118	0.132
Yunan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Shaanxi	0.348	0.356	0.339	0.325	0.302	0.283	0.264	0.275	0.276	0.256	0.248	0.242
Gansu	0.324	0.324	0.312	0.306	0.298	0.267	0.256	0.248	0.242	0.232	0.237	0.237
Qinghai	0.262	0.225	0.226	0.220	0.223	0.216	0.198	0.176	0.165	0.158	0.163	0.154
Ningxia	0.406	0.398	0.385	0.395	0.346	0.293	0.254	0.246	0.253	0.234	0.236	0.250
Xinjiang	0.298	0.293	0.284	0.269	0.256	0.243	0.227	0.206	0.207	0.198	0.192	0.190

Data are calculated according to Equation (1) by using collected data

Environmental efficiency of each province from 2001 to 2012 Values of environmental efficiency are generally small with a downtrend for most provinces. Shanghai, Beijing and Tianjin's efficiency value has always been 1, making them on the frontier. At the same time, they have become the standard of the other provinces to measure efficiency value, The value for Fujian have been 1 for 6 years from 2001–2006; but it fail to reach the frontier later though it is relatively big. This change is caused by too much waste water discharge according to analysis of redundancy rate, Efficiency values are the smallest for Guizhou, Jiangxi, Qinghai, Sichuan and Chongqing with average value smaller than 0.2, far from the frontier. Referring to Beijing, Tianjin, Shanghai, Yunnan and Liaoning, polluting emission in Guizhou, Jiangxi, Qinghai and Sichuan can be reduced by about 80% with the same input and output. Analysis shows that in front of the value, there is a huge difference among various provinces, and the potential of reducing polluting emission is bigger for provinces with smaller values. For Zhejiang, Shandong, Jiangsu and Guangdong, Because of the high energy consumption and pollution emissions, their efficiency values decrease obviously though these provinces are economically developed areas. One more point should be stated clearly. Efficiency value measured in our article is just a comparative index, meaning provinces on the producing frontier enjoy some advantages. There absolutely exists room of reducing polluting emission for them. In the exploration and the application of environmental protection technology, China is still a long way to go. High efficiency value of

provinces has an obligation to improve environmental protection and reduce pollution emissions.

5 Conclusions

According to the above analysis, we can give some advice on improving the efficiency of China's environmental protection: (1) Continue to give top priority to the east, using The advantage of the developed economy, energy conservation and emissions reduction to change the industrial structure, To improve the technology of industry, productivity, and reasonable development of the tertiary industry, To improve the ratio of the third industry in the national economy, to build a resource conservation and environmentally friendly society. (2) According to regions' different levels, making out the energy conservation and emissions reduction goals and policies, focus on the central and western regions, where there is more room to improve. With a low environment efficiency to balance the regional difference. Since the west development policy, the economic development of the western is good, need to increase the area in terms of production technology research, and communicate with other regional cooperation. And at the same time, Under the background of the rapid development in the Midwest, central also need to accelerate the industrial innovation, optimize the industrial structure and the allocation of resources, improve efficiency, so as to accelerate the economic development, realize the central and western regions of the strategic significance of ecological

environment for China's sustainable development. (3) To strengthen environmental control, according to the domestic and international situation, making energy conservation and emission reduction policy, coordinate relationship between economic development and environmental protection, to guide the local government to improve environmental efficiency. (4) Accelerate the

development of science and technology, especially in the energy saving of high technology innovation to give more policy support, ensure that the new technology faster and more widely applied to each region, industry, and the production process, as well as ensuring the environmental efficiency steadily increases with rapid economic development.

References

[1] Cerutti A K, Bruun S, Beccaro G L, Bounous G 2011 *Journal of Environmental Management* **92** (10) 2277-86
 [2] Fare R, Grosskopf S, Lovell C A K, Pasurka C 1989 *The Review of Economics and Statistics* **71** 90-8
 [3] Zhou P, Ang B W, Poh K L 2008 *European Journal of Operational Research* **189** 1-18
 [4] Wang K, Yu S W, Zhang W 2013 *Mathematical and Computer Modelling* **58**(5-6) 1117-27
 [5] Yeh T L, Chen T Y, Lai P Y 2010 *Energy Policy* **38** 2386-94
 [6] Yang L, Wang K L 2013 *Mathematical and Computer Modelling* **58**(5-6) 1074-83
 [7] Wang Z H, Zeng H L, Wei Y M, Zhang Y X 2012 *Applied Energy* **97** 115-23
 [8] Zhou P, Poh K L, Ang B W 2007 *European Journal of Operational Research* **178** 1-9
 [9] Shi G M, Bi J, Wang J N 2010 *Energy Policy* **38** 6172-9
 [10] Barba-Gutiérrez Y, Adenso-Díaz B, Lozano S 2009 *Environmental Modeling & Assessment* **14** 439-47
 [11] Azadeh A, Ghaderi S F, Omrani H, Eivazy H 2009 *Energy Policy* **37** 2605-18
 [12] Mukherjee K 2010 *European Journal of Operation Research* **201** 933-41
 [13] Riccardi R, Oggioni G, Toninelli R 2012 *Energy Policy* **44** 140-52
 [14] Chiu Y, Wu M 2010 *Polish Journal of Environmental Studies* **19** 1159-69
 [15] Zhao Y, Zhang Y, Guang J 2013 *Nature Environment and Pollution Technology* **12** (4) 569-75

Authors	
	<p>HaiLei Zou, born in July, 1978, PuTuo County, ZheJiang Province, P.R. China</p> <p>Current position, grades: Lecturer of School of Science, China JiLiang University, China. University studies: M.Sc. from Chong Qing Normal University in China. Scientific interest: operational research, financial mathematics Publications: more than 5 papers published in various journals. Experience: teaching experience of 10 years, 2 scientific research projects.</p>
	<p>Cheng Wang, born in June, 1980, He County, AnHui Province, P.R. China</p> <p>Current position, grades: Lecturer of School of Science, China JiLiang University, China. University studies: M.Sc. from Zhejiang University in China. Scientific interest: stochastic process, financial mathematics Publications: more than 6 papers published in various journals. Experience: teaching experience of 10 years, 2 scientific research projects.</p>

Correlation model analysis on the land price fluctuations in Beijing and Tianjin City in China

Zhaoxia Si^{1, 2}, Shaoliang Zhang^{1*}, Ningli Chen²

¹*School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, 221116, China*

²*Henan Polytechnic University, Jiaozuo, 454000, China*

Received 8 January 2014, www.tsi.lv

Abstract

According to the indexes of the city land price, the co-integration analysis, the Granger causality test, the impulse response function and the variance decomposition method were used in this paper to analyse the correlation model of the land price fluctuations in Beijing and Tianjin city in China. The results showed that the land prices in Beijing and Tianjin city had a long-term co-integration relationship. The Granger test model showed that the land prices in Beijing and Tianjin city was in line with positive correlation bidirectional causality. In addition, the raise of 1% land price in Beijing city caused a raise of 0.96% in Tianjin city. Conversely, a raise of 1% in Tianjin city caused a raise of 1.03% in Beijing city. By comparing the mutual influence degree, the land price fluctuations in Beijing city had a greater influence on that in Tianjin city.

Keywords: Land price fluctuation model, Co-integration test, Granger causality model, Impulse response function

1 Introduction

The continuously rising about land price in large and medium-sized cities in China causes public concerns. Urban land prices represent regional markets with obvious heterogeneity. However, such prices do not exclude the spatial relevance about the land price of regional cities in China. The research about the regional relevance model of urban land prices plays an important role in helping city governments to implement regulatory policies in housing markets.

Abundant studies have been done to regarding the internal relationship model between regional housing prices and price conduction. In 1982, Mcavinchey and MacLennan studied the expansion rate of the regional housing price in the United Kingdom [1]. In the 1990s, according to Holmans' study, the changes in house prices in the southeast region of the UK lead to changes in the house prices in other regions and this phenomenon, in terms of price fluctuations, is referred to in academic circles as the "Ripple Effect" [2]. Currently, a large number of studies support this viewpoint. Giussani and Hadjimatheou studied the relative changes of housing prices in the southeast and the northwest regions in UK [3]. By the Engle-Granger test method and the Johansen co-integration test, MacDonald and Taylor studied the co-integration relationship of regional housing prices in the UK [4]. Alexander and Barrow (1994) validated the existing co-integration relationships in regional housing prices in the UK [5]. The price of urban housing in the southeast region of the UK follows the Granger theory of fluctuation of urban housing prices, with the ripple effect

spreading to the southern region. Based on a spatial econometrics and co-integration theory, Meen studied the regional housing prices in the UK, which indicates that the rate of price fluctuations will stabilize in the long term even though housing prices fluctuate in the short term, [6]. Cook pointed that regional housing prices in the UK are convergent, and when the housing prices in the southern region of the UK fluctuated, the equilibrium state would be returned quickly [7, 8], which in turn verified the existence of a stable long-term relationship in regional housing prices. By the method of unit root tests of panel data, Holmes studied the variations in regional housing prices in the UK and found that the most regional housing prices were convergent [9]. Based on the changes in regional housing prices in the northeast and mid-east regions of America, Pollakowski and Ray found that the housing prices in adjacent metropolitan areas were conductive [10]. Gupta and Miller used the ripple effect theory to analyse the relationship of housing prices in some American cities through time-series data [11]. Oikarinen and Liu investigated the changes in regional housing prices in the interval between Finland and Australia, respectively [12, 13]. At present, no consensus has been reached on the reasons of the regional price fluctuation. The most notable research is the one conducted by Meen [14]. According to existing Literatures, the family migration, wealth transfer, transaction and search costs, spatial arbitrage and exogenous shock with spatial time-lag effect should be included to resolve this phenomenon.

In China, scholars pay more attention to the influence factors on urban and national housing prices. However,

* *Corresponding author* e-mail: sizhaoxia@hpu.edu.cn

less attention was paid to the relevance of housing price fluctuations model between different cities. Only a few scholars have successively did research into the ripple effect of housing prices based on some related foreign studies. Liang et al. [15] compared the fluctuations of housing prices in various cities and found obvious regional imbalances. In order to explore the formative reasons, they further studied the differences in the regional fluctuation of housing prices by constructing a model of panel data. Hong studied the linkage of real estate prices about various Chinese cities [16]. With the methods of the Johansen co-integration regression test and the multi-variable Granger test of causality and impulse response function (IRF), Wang et al. drew on the ripple theory and analysed the interactive relationship model between housing prices in the cities of five major Chinese regional markets [17]. These studies indicated that the significant differences exist in short-term fluctuations in urban housing prices, while a stable competitive relationship could be observed in the longer term fluctuations.

Based on the studies conducted by scholars in China and other countries, it could be argued that the scholars in other Countries have done relatively extensive studies on the fluctuations of regional housing prices. Although the time for these researches was relatively short, the contexts of the foreign studies cover a wide range of factors and focus more on the theoretical exploration. Scholars in other countries adopted an econometric method to quantitatively analyse the relevance of the fluctuations of regional prices.

Based on the research methods from scholars in other countries, and taking the actual situation in China into consideration, scholars in China conducted their studies about the fluctuation of regional urban housing prices. The results verified the existence of the ripple effect, and the reasons were preliminarily discussed. According to the existing references, the studies specializing in the relevance model of land prices in China's mega-cities are rare. In this paper, the land prices in the mega-cities of Beijing and Tianjin were selected as the main research objects. The relevance model study about land price for different cities is of great significance. For China, the acquisition of a linkage mechanism of the land prices between cities can provide a reference point for governments to make decisions to establish differentiated regulatory policies.

2 Data Resources and Study Methods

Urban housing price can reflect the urban land prices. In this paper, the aggregative indexes of urban housing prices in Beijing and Tianjin are selected for analysis. Selection for price index can reduce heteroscedasticity among data, increase the stationarity of series and eliminate the effect on the results. The data obtained from monthly monitoring data was selected from 2005 to August of 2012, which is from the China Index Research

Center, the Beijing Bureau of Statistics and the Real Estate Information NET.

The fluctuation of urban land prices will bring corresponding changes to the urban land prices of adjacent regions. The key for studying the relevance of the fluctuation of regional urban land prices is to discuss how and why the regional land prices fluctuate. By using Eview 6.0 analysis software and adopting co-integration analysis, the Granger causality test, the impulse response function and the variance decomposition method, the relevance mechanism on the fluctuation of urban land prices was analysed and the model is built as follows [18-25]:

$$HPB_t = C + \alpha \times HPA_t + \varepsilon_t, \quad (1)$$

where, HPB_t represents the land price index of B city, i.e., the affected city; HPA_t represents the land price index of A city, i.e., the land price of this city changes first; C is a constant; α is the coefficient; ε_t is the random variable.

3 Empirical Analysis

3.1 CORRELATION MODEL ANALYSIS ON THE URBAN LAND PRICES BETWEEN BEIJING AND TIANJIN CITY

From Figure 1 and Figure 2 the change trends about the urban land prices of Beijing and Tianjin city show relatively high consistency.

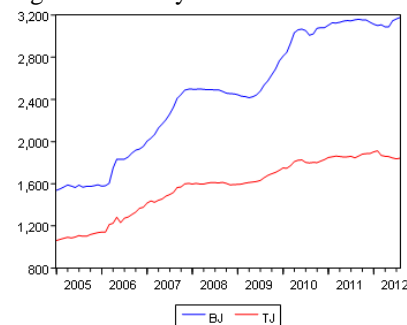


FIGURE 1 Sequence correlation diagram of land prices in Beijing and Tianjin

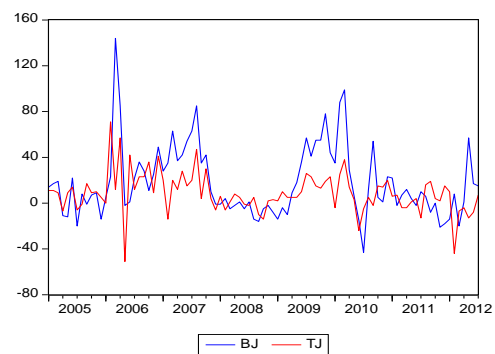


FIGURE 2 Changing sequence correlation diagram of land prices in Beijing and Tianjin

The changes amplitude of the urban land prices in Beijing city is higher than that in Tianjin city. Therefore, it can be concluded that the changes of urban land prices in Beijing and Tianjin embody a relevance model.

3.2 DATA STATIONARITY TEST

A co-integration test requires the same integration order for different variables. In order to apply the co-integration theory to test the long-term relationship between the variables of the urban land prices in Beijing and Tianjin city, the stationarity of the variables is tested firstly. The “stationarity of time series” generally refers to the fact

that the statistical rule of time series changes little along with time. In other words, the process of generating random variable does not change along with time. The ADF model is adopted to conduct a unit root test on the time series of land prices for the two cities, and the results are shown in Table 1. The results indicate that the time series of an urban price index are not stationary, but by ordinary differential analysis, each time series is stationary and can be seen as a first order integrated series I(1). Based on the unit root test, the co-integration relationship between the time series of the urban land prices in both cities could be tested by I (1).

TABLE 1 Unit root test on the time series of land prices in Beijing and Tianjin city

Variate	Test Type (c,t,n)	ADF statistical magnitude	1% critical value	5% critical value	Inspection results
BJ	c,0,0	-1.0798	-3.5047	-2.8939	nonstationary
TJ	c,0,0	-2.4314	-3.5039	-2.8936	nonstationary
D(BJ)	c,0,1	-4.7010	-3.5047	-2.8939	steady
D(TJ)	c,0,1	-4.8275	-3.5056	-2.8943	steady

Notes: D(BJ) and D(TJ) are the original sequence of first order difference. (c,t,n) are respectively constant term, trends, and lag order number.

3.3 JOGANSEN CO-INTEGRATION TEST

The main principle of the Johansen test is that if the co-integration vector matrix contains a non-zero characteristic root and how many non-zero characteristic roots are contained. The existence of non-zero characteristic roots can indicate that there is a co-integration relationship among various time series. The Trace Statistic and Max-Eigen Statistic could be applied

to conduct this test. The test shows that the system has no linear trend, and the co-integration equation has intercept terms but no trend terms. The results in Table 2 indicate that under the level of 5%, the Trace Statistic and Max-Eigen Statistic exists a co-integration relationship, and a long-term equilibrium relationship exists between the urban land prices in Beijing and Tianjin city.

TABLE 2 Johansen test between urban land prices in Beijing and Tianjin city

Original hypothesis cointegration equation	Eigenvalue	Trace Test	Trace Test C.V.	Maximum Eigenvalue	Maximum Eigenvalue C.V.
None *	0.16269	21.9879	20.2618	15.9810	15.8921
At most 1	0.06456	6.0069	9.1645	6.0069	9.1645

Notes: “*” is that refused to the original assumption under the 0.05% significant level.

3.4 GRANGER CAUSALITY TEST

The co-integration test illustrates that there is a long-term equilibrium among variables. However, further testing needs to be done to ensure if the causality exists. Granger proposed a simple test program, which is habitually called as the Granger causality test. Granger used the relationship between time series to define the causality, and proposed an econometrical definition of causality as follows. If y could be used to estimate another quantity x, the y would be regarded as being able to cause x. The test model for this theory is as follows:

$$Y_t = \sum_{i=1}^m \alpha_i X_{t-i} + \sum_{i=1}^m \beta_i Y_{t-i} + \mu_{1t} \tag{2}$$

$$X_t = \sum_{i=1}^m \lambda_i Y_{t-i} + \sum_{i=1}^m \phi_i X_{t-i} + \mu_{2t} \tag{3}$$

By constructing F statistics, the null hypothesis of F test without Granger causality is used.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = L = \beta_i = 0, i = 1, 2.$$

The test results could be obtained by comparing the F statistics and the size of critical value. If F is greater than the critical value, he null hypothesis that “Y is not the cause of X” should be rejected. In other words, Y is the cause of X. If F is smaller than the critical value, the null hypothesis could not be rejected, which means that X is not the cause of Y.

According to the above ADF test results, the first order differences of the urban land price indexes in Beijing and Tianjin city are both stationary series, therefore, the test of Granger causality can be conducted. As far as the Ganger causality test, the lagged difference shall be set artificially. The lagged order could exert a significant influence on the results, and a lower order would result in some of the important variables being

ignored. A higher order would reduce the degree of freedom of the model, and a larger standard deviation occurs to parameter estimation, and thus, the level of precision is decreased. In this paper, the determination of lag period establishment was carried out according to the

lag length criteria. From Table 3, the minimum lag period provided by each of the five evaluation indexes should be represented by “*”. According to AIC and SC minimum criteria, the optimal lag period is selected as the second period.

TABLE 3 Criteria to determine VAR lag length and appropriate lag order number

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1147.113	NA	4.22e+08	25.53584	25.59139	25.55824
1	-803.6885	663.9536	223546.3	17.99308	18.15973	18.06028
2	-779.1029	46.43956*	141505.6*	17.53562*	17.81338*	17.64763*

The diffusing directions of the two variable land prices in Beijing and Tianjin city are investigated, and the Granger causality test is applied. The results in Table 4 show that a bi-directional causality exists between the

land prices in Beijing and Tianjin city. In other words, changes in the land prices in Tianjin cause changes in the land prices in Beijing, and vice versa.

TABLE 4 Granger causality test

Null Hypothesis:	F-Statistic	Prob.	Results
D(TJ) does not Granger Cause D(BJ)	4.2753	0.0170	Accept
D(BJ) does not Granger Cause D(TJ)	7.8177	0.0007	Accept

3.5 ANALYSIS ON THE IMPULSE RESPONSE AND CONTRIBUTION

The above research analysed the long-term equilibrium relationship between land prices in Beijing and Tianjin city was, and the causality between them was also discovered. To further understand the change rule within a certain period and the impact on their long-term equilibrium relationship, both impulse response analysis

and variance decomposition based on co-integration analysis are required.

3.5.1 Impulse response analysis

Impulse response analysis is generally conducted by observing the graph of impulse response function. This function depicts the response to great or small error changes.

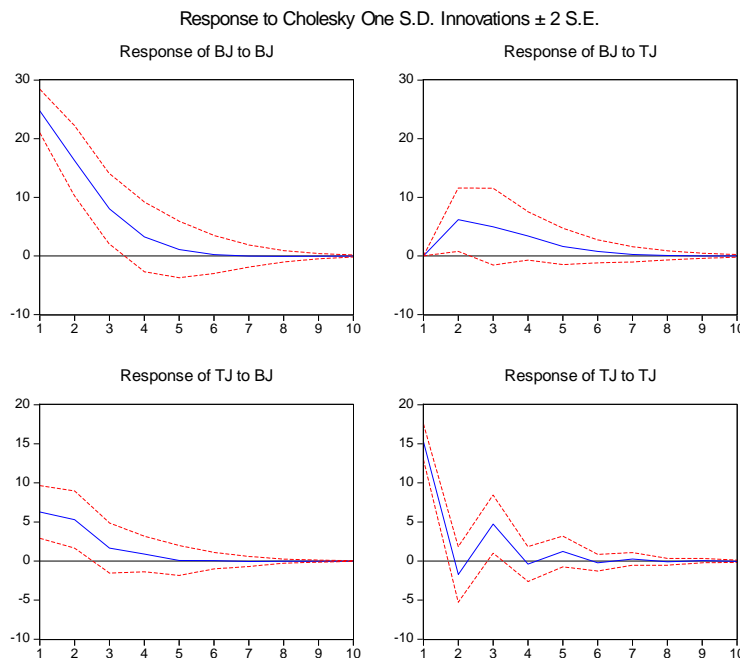


FIGURE 3 Impulse response function of land prices in Beijing and Tianjin

Supposing the random disturbance term is impacted by a standard deviation, the variables will always be affected both currently and for several years after the initial deviation, i.e., the influence of change or the

impact of each endogenous variable on itself and other endogenous variables. Figure 3 indicates the investigation of the response paths of the two variables – land prices in Beijing and Tianjin - within 10 units of time for the

impact on price when the given land price impact is 1%. The solid lines represent the change traces of BJ and TJ after being impacted; the dotted lines represent the values of BJ and TJ adding and subtracting twice the standard deviation, indicating the maximum range possibly influenced by BJ and TJ after being impacted, i.e., the attained upper and lower limits.

Figure 3 shows that after a random disturbance on one of its standard deviations, the land price in Beijing shows an immediate strong reaction by rising about 25 grades of intensity. Its impulse curve reveals a decline in the first four months, then gradually climbs back, basically keeping a position near to zero and then practically disappearing in the seventh month. All these influences are positive. After the land price in Beijing affected by a standard deviation, the land price in Tianjin is also impacted, which is not a direct price rise in the same month, but a maximum influence in the next month with an increase in intensity of five grades? Then the increase rate gradually falls back and practically disappears in the eighth month, and also, all these influences are positive. This indicates that if the land price in Beijing rises due to the influence of certain factors, the stimulatory effect will not last too long. This in turn indicates that when the land price itself undergoes a smooth development, the residents' purchasing power will be relatively stable. Even the factors or events prompt centralized consumption by residents over a certain period of time,

an increase in long-term consumption will not follow.

After a random disturbance caused by one of its standard deviations, the land price in Tianjin also shows a relatively strong reaction with a rise of 15 grades of intensity. Its impulse curve presents a diminishing waveform with two months as a cycle, and practically disappears in the eighth month. Tianjin imposed a smaller level of disturbance on Beijing, and such disturbance gradually descends and practically disappears in the fifth month.

3.5.2 Contribution analysis

When a certain variable of the system suffers from one unit of impact, the degree of the interaction between variables could be reflected by a variance in the percentage of prediction errors of the variables, which is the variance decomposition. This variance decomposition could offer relatively important random information. To be specific, the change of each endogenous variable is decomposed into components associated with various random disturbance terms, according to the cause of the change, so that the relative importance of each innovation to the endogenous variables of the model can be understood. The variance decomposition analysis is required to investigate the factors influencing the fluctuation of land prices in Beijing and Tianjin.

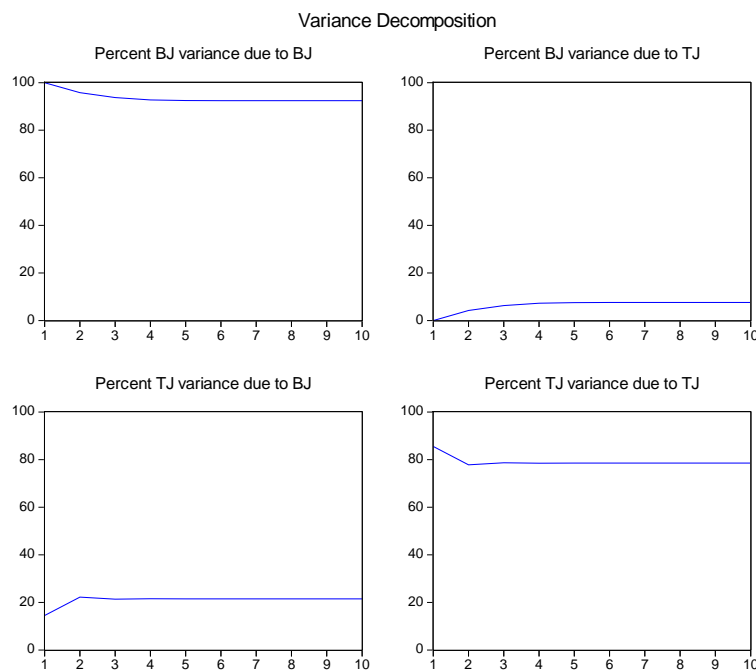


FIGURE 4 Variance decomposition of land prices in Beijing and Tianjin

According to Figure 4, with an initial increase in the in Beijing gradually decreases, while the influence on the land price changes in Tianjin gradually increases. Up to the fourth predictive period, the information shares of the land prices in Beijing and Tianjin tend to be relatively stable. The information shares of the land prices in

predictive period, the influence of the land price changes Beijing and Tianjin are 92.51% and 7.49%, respectively. In the same way, the influence of the land price in Tianjin gradually decreases initially, while the influence on the land price in Beijing gradually increases. The trend is not stabilizing until the second predictive period,

when the information shares of Beijing and Tianjin are 80.00% and 20.00%, respectively. According to the contribution analysis, the influencing share of the land price change in Beijing on the land price fluctuation in Tianjin only accounts for 7.49%, while the influencing share of the land price change in Tianjin on the land price fluctuation in Beijing reaches 20.00 %. These figures suggest that during an interaction between land price changes in Beijing and Tianjin, the degree of any land price fluctuation in Tianjin caused by a land price fluctuation in Beijing will be lower than the degree of any land price fluctuation in Beijing caused by a land price fluctuation in Tianjin.

3.6 MODEL ANALYSIS ON LAND PRICE FLUCTUATION OF BEIJING AND TIANJIN CITY

Eview6.0 analysis software is used to build the analysis model of the land prices in Beijing and Tianjin city, and the results obtained by the model are as follows:

$$HP_{TJ} = C + \beta \times HP_{BJ} + \varepsilon_t, \tag{4}$$

$$HP_{BJ} = C + \delta \times HP_{TJ} + \varepsilon_t. \tag{5}$$

TABLE 5 Coefficient and conduction direction of land prices in Beijing and Tianjin

Conduction factor	Beijing → Tianjin		Tianjin → Beijing	
	coefficient	Adjusted R ²	coefficient	Adjusted R ²
HP	0.9579 (β) (30.0836)	0.9858	1.0292 (δ) (30.0836)	0.9858

According to Table 5, the high regression coefficient of the model is 0.98, which indicates a good-fitting result of the model. When the land price in Beijing city rises by 1%, the land price in Tianjin city will rise by 0.96%. When the land price in Tianjin rises by 1%, the land price in Beijing will rise by 1.03%. These are consistent with the variance decomposition results.

4 Conclusions

The price relevance model of the land price indexes from 2005 to 2012 in Beijing and Tianjin city was studied, and the fluctuation model of the land prices of the two cities was investigated by the methods of the co-integration test, the causality test, the impulse response function and the variance decomposition.

The co-integration test result indicates that a long-term co-integration relationship exists between the land prices in Beijing and Tianjin city. Even the land prices are externally impacted, they will soon return to a state of equilibrium.

The Granger causality test result indicates that the land prices in Beijing and Tianjin city exist bi-directional causality. The influence on land prices of the two cities is mutual.

According to the quantitative analysis on the fluctuation of land prices in Beijing and Tianjin city, a positive correlation relationship model exists. When the land price in Beijing rises by 1%, the land price in Tianjin will rise by 0.96%; when the land price in Tianjin rises by 1%, that increase will cause the land price in Beijing to rise by 1.03%.

According to the contribution analysis, the land price changes in Beijing city can cause only 7.49% fluctuations in Tianjin city. However, the land price changes in Tianjin can cause 20% fluctuations in Beijing city. During the interaction of land price changes in Beijing and Tianjin, the degree of any land price fluctuation in Tianjin caused by a land price fluctuation in Beijing is lower than that in Beijing caused by a land price fluctuation in Tianjin city. This is consistent with the quantitative analysis on the land price fluctuation of the two cities.

Acknowledgments

The authors are grateful for the financial support provided by the National Foundation of China (No. 40971074).

References

[1] McAviney I D, Maclennan D 1982 *Urban Studies* 19(1) 43-57
 [1] Holmans A E 1994 *Economic Modelling* 11(2) 157-99
 [2] Giussani B, Hadjimatheou G 1991 *Papers in Regional Science* 70(2) 201-19
 [3] MacDonald R, Taylor M P 1993 *Scottish Journal of Political Economy* 40(1) 43-55
 [4] Alexander C, Barrow M 1994 *Urban Studies* 31(10) 1667-89
 [5] Meen G 1996 *Housing Studies* 11(3) 345-72
 [6] Cook S 2003 *Urban Studies* 40(11) 2285-94
 [7] Cook S 2005 *International Review of Applied Economics* 19(1) 107-18
 [8] Holmes M J 2007 *Journal of Economic and Social Research* 9(1) 1-17
 [9] Pollakowski H O, Ray T S 1997 *Journal of Housing Research* 8(1) 107-24
 [10] Gupta R, Miller S M 2012 *The Annals of Regional Science* 48(3) 763-82
 [11] Oikarinen E 2004 *Journal of Housing Research* 15(1) 3-28
 [12] Liu C L, Luo Z Q, Ma L, Picken D 2008 *International Journal of Strategic Property Management* 12(4) 237-50
 [13] Meen G 1999 *Housing Studies* 14(6) 733-53
 [14] Liang Y F, Gao T M 2007 *Economic Research Journal* (8) 133-42 (in Chinese)
 [15] Hong T, Xi B, Gao B 2007 *Statistical Research* 24(8) 64-7 (in Chinese)
 [16] Wang S T, Yang Z, Liu H Y 2008 *Research on Financial and Economic Issues* (6) 122-7 (in Chinese)
 [17] Zhang X T 2009 *Use guidelines and examples of Eviews* China Machine Press: Beijing chapter 7 (in Chinese)

- [18]Gao T M 2006 *Econometric analysis method and modelling-reviews application and instance* Qinghua University Press: Beijing chapter 4-5 (in Chinese)
- [19]Zhang Q Y 2008 *China Land Science* 22(12) 68-76 (in Chinese)
- [20]Zhong W 2010 *Study on the housing price diffusion of pole cities in the pearl river delta*, Doctoral Dissertation of Huazhong University of Science and Technology, China (in Chinese)
- [21]Gao J 2010 *The research of Chinese real estate price fluctuation* Doctoral Dissertation of Beijing Technology and Business University, China (in Chinese)
- [22]Zhao T T 2012 *Transmit characteristics and influential factors of housing price in Changzhutan urban agglomeration* Doctoral Dissertation of Hunan Normal University, China (in Chinese)
- [23]Ai J G, Ding L Y, He S B 2008 *Urban Studies* 15(1) 77-83 (in Chinese)
- [24]Zang B, Yang Q Y, Zhou T 2011 *China Land Science* 25(5)79-84 (in Chinese)

Authors	
	<p>Zhaoxia Si, born in March, 1980, Gaoping County, Shanxi Province, P.R. China</p> <p>Current position, grades: PhD student of School of Environment and Spatial Informatics, China University of Mining and Technology, China. Lecturer of Henan Polytechnic University, China.</p> <p>University studies: B.Sc. from Taiyuan Normal University in China. M.Sc. from Sun Yat-Sen University in China.</p> <p>Scientific interest: Land economics, Land monitoring and valuation</p> <p>Publications: 3 papers published in various journals.</p> <p>Experience: teaching experience of 8 years, 3 scientific research projects.</p>
	<p>Shaoliang Zhang, born in March, 1968, Anqing County, Anhui Province, P.R. China</p> <p>Current position, grades: Associate Professor of School of Environment and Spatial Informatics, China University of Mining and Technology, China.</p> <p>University studies: B.Sc. and M.Sc. from China University of Mining and Technology</p> <p>Scientific interest: Evaluation of land resources, Mine ecological monitoring and evaluation</p> <p>Publications: more than 40 papers published in various journals.</p> <p>Experience: teaching experience of 24years, 20 scientific research projects.</p>
	<p>Ningli Chen, born in April, 1988, Zhoukou County, Henan Province, P.R. China</p> <p>Current position, grades: Master graduate student of Henan Polytechnic University, China.</p> <p>University studies: M.Sc. from Xinyang Normal University in China.</p> <p>Scientific interest: Land monitoring, ecological evaluation</p> <p>Publications: two papers published in various journals.</p> <p>Experience: teaching experience of 2 years, 2 scientific research projects.</p>

Dynamic evaluation and simulation of variant-driven complexity costs in multi-echelon automotive supply chains

Haiyan Yi¹, Dianjun Fang^{2*}

¹*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu City, Sichuan Province, China, 610031*

²*Chinese-German School for Postgraduate Studies, Tongji University, Shanghai City, China, 200092*

Received 9 March 2014, www.tsi.lv

Abstract

The proliferation of product variety driven by part variants imposes a great impact on the costs, performance and ecological burden of automotive supply chains. Costs are always the first aspect to be considered by manufacturers. This paper creates a multi-echelon automotive supply chain in the light of the process model and improves the KPI of low costs to evaluate the variant-driven complexity costs. Then, the evaluation models of different complexity costs are constructed. An automotive supply chain scenario model is constructed with the OTD-NET and the simulation results are analysed. The study of this paper will be helpful to automotive manufacturing in seeking to optimize the number and range of product variants.

Keywords: Automotive Supply Chain, Product Variety, Complexity Costs, OTD-NET

1 Introduction

Part variants are the main factors of product variety in automotive industry, as discussed by Lechner [1] and Klingebiel and Yi [2], so it is very important to evaluate their influence on automotive supply chains. The evaluation of supply chains can be carried out through economy, performance and ecology, as discussed by Klingebiel [3], in which the economy factor is always considered first since achieving profits is the primary objective for business. Therefore, reducing costs or increasing profits are crucial for evaluating or choosing a strategy.

Many researchers have focused on the impact of the proliferation of product variety on the costs or profits of automotive supply chains. As pointed out by Ramdas [4], simply increasing variety may not guarantee growth in long-term profits, and may even worsen a firm's competitiveness. Fisher et al. [5] and Schleich et al. [6] argued that a large product variety may increase overhead costs, resources costs, production costs and inventory costs. As variety increases, the plant may have to pay a higher cost in the form of fixed investments in variety handling systems, which is much higher than the variable cost of variety as discussed by Fisher and Ittner [7, 8] and Ittner and MacDuffie [9]. Jagersma [11] also concluded that complexity tends to increase the fixed costs of conducting business. As discussed elsewhere [1, 4, 11–20], other complexity costs that may be caused by an increase of product variety include retailers' costs, set-up costs, operational cost, overhead costs, raw material costs, equipment costs, management costs, space costs, etc. In general, Schoeller [21] claims that costs associated

with excessive complexity become the main cost factors of supply chains, which can result in a considerable reduction of profit (as discussed by Berry and Cooper [22] and Kersten et al. [15]).

Therefore, it is necessary for automotive companies to discover the most suitable degree of product variety to control the increase of costs, as discussed by Benjaafar et al. [17], Scavarda et al. [23], Schleich et al. [6] and Thonemann and Bradley [11]. Scavarda et al. [23] also show that companies should evaluate how the costs can be influenced by variety in order to achieve this aim, and that variety in the automotive industry should be split between factory-fitted (variants) and dealer-fitted options. Nevertheless, Lechner [1] and Yi and Klingebiel [2] consider there is still a lack of research, especially on variants, as well as a lack of dynamic evaluation and simulation of this type of impact.

This paper aims to evaluate the impact of the growth of part variants on logistics costs in the automotive supply chain through the application of the OTD-NET simulator. The next section constructs a four-echelon automotive supply chain and examines the key performance indicators (KPIs) of logistics costs. We then build the models for evaluating different complexity costs induced by variant-driven variety, before the four-echelon supply chain is simulated and the results are analysed. The paper concludes and discusses future studies.

2 KPIs of logistics costs

The KPI system should be specified before the evaluation of complexity costs. This section constructs a four-echelon automotive supply chain as an area for study

* *Corresponding author* e-mail: d.fang@gmx.de

before generalizing about the KPIs in the light of the process chain model.

2.1 A FOUR-ECHELON AUTOMOTIVE SUPPLY CHAIN MODEL

The four-echelon automotive supply chain from suppliers to dealers that is studied and simulated in this paper is shown in Figure 1. Each supplier offers one part type, which may consist of several part variants, for the plant. The variety of finished cars is sold through a couple of dealers or importers who account for actual rates, respectively.

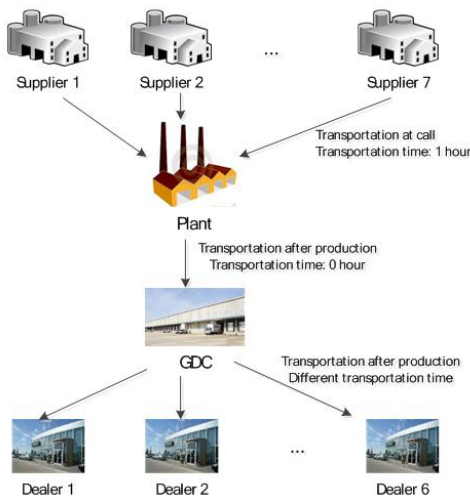


FIGURE 1 General structure of the four-echelon automotive supply chain model

In this supply chain, the suppliers are co-located with the plant in the same city in Germany. The transportation time from any supplier to the plant is one hour. As the product structure of the automotive industry is too complicated to simulate all the part variants, only seven part types are built in this scenario model to compose the finished product: “PC car”. These seven part types are body sheets, paint, engines, transmissions, seats, door and locks, wheels and tires. All suppliers are build-to-order (BTO) suppliers, except for body sheets and paint, which are build-to-stock (BTS). Only the BTS suppliers keep a minimum stock for each component in the plant’s warehouse, which is equal to the demands of three hours supply. There are four workshops in the plant: press, welding, paint and assembly. The labour time for all the workshops in the plant is 21 hours, from Monday to Friday, with three shifts, and the planned output of the plant is 50 per hour. The finished cars are driven to the global distribution centre (GDC) once they have rolled off the production line. The GDC is located in the plant area, so the transportation time from plant to GDC is considered to be zero. Four of the six dealers are in Germany and the others are in neighbouring countries. The transportation time for finished cars from GDC to each dealer varies from six to 24 hours, except for the ten minutes of one dealer lot that is co-located in the GDC. It is assumed that all customers belong to the same

customer type, and want to receive their products as soon as possible and do not accept any deviations of the product. The dealers in different markets gather orders from the customers and send these orders to the plant every Monday. The GDC transport the finished cars once they have satisfied the minimum transport units.

The logistics process in the supply chain can be divided into three processes centring on the plant: inbound logistics; in-plant logistics; and distribution logistics. Inbound logistics refer to parts transported from suppliers to plant and parts stocked in the warehouse; in-plant logistics include producing processes in the plant and finished products stocked in the GDC; distribution logistics are those cars distributed from the GDC to dealer lots and cars stocked in dealer lots. Each logistics process consists of several processes, structures and resources according to the process chain model, which is proposed by Kuhn [24]. The process chain model visualizes and abstracts the material and information flows in each echelon of supply chains, so that the steps and their relationships can be grasped clearly (as studied by Kuhn and Hellingrath [25]). With an increase of variants, the structures should be changed to deal with this, and additional processes and resources are required. The KPI of evaluation should reflect the costs resulting from these changes.

2.2 KPI OF EVALUATION OF COMPLEXITY COSTS

The general target of low costs for the supply chain can be described with respect to low process costs, low inventory costs and low resources costs, as discussed by Benjaarfah et al. [17], Lechner [1], Lechner [1], Thonemann and Bradley [11] and VDI [26]. Figure 2 summarizes the indicators with respect to these costs.

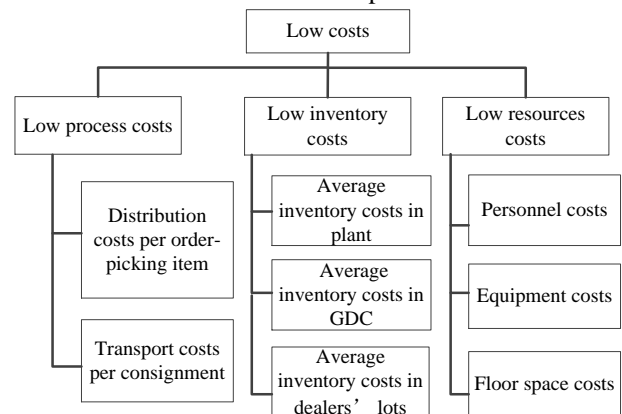


FIGURE 2 KPI of Low Costs to Evaluate the Variant-driven Product Variety

In order to improve the accountability of these indicators, they are analysed in combination with the logistics process discussed above. Table 1 shows the indicators with respect to logistics processes and functions.

The low process costs are measured by two indicators: distribution costs per order-picking item and

transport costs per consignment. For the purpose of achieving coherence between the description of supply chains, the definition in OTD-NET and the KPI system, the distribution costs indicator is adopted to evaluate the process costs of delivery goods from GDC to dealer lots, while the transport costs indicator evaluates the process costs of transport parts from suppliers to plant.

The second sub-objective, which of low inventory costs, aims for a quick turnover of inventories in plant, GDC and dealer lots. The third sub-objective, low resource costs, can be evaluated from three indicators: personnel costs; equipment costs; and floor space costs. The personnel costs and equipment costs during the production logistics are both proportional to the process time of production, as demonstrated by Lechner [1]. The

floor space costs are the costs of space in workshop, warehouse, etc.

It is obvious that variety is a “double-edged sword”, as shown by Karmarkar [27] and Shapiro [28]. An automotive supply chain should plan the suitable number of variety and variants for their products. Some approaches should evaluate the variant-driven complexity costs so that cost–benefit analysis of variant-drive product variety in the automotive supply network can be carried out. Here, the variant-driven complexity costs refer to the difference in logistics costs in the automotive supply network between a scenario with variants and a zero-based scenario. The next section introduces the evaluation approaches apart from the zero-based approach, and constructs the evaluation models based on such approaches.

TABLE 1 Function costs and process costs

		Logistics/functional costs (CTF)		
		Process costs (CP)	Inventory costs (CI)	Resources costs (CR)
Logistics processes costs (CTP)	Inbound logistics costs (CP ^B)	Transport costs	Inventory costs of parts in plant	Personnel costs; equipment costs; floor space costs
	In-plant logistics costs (CP ^P)		Inventory costs of unfinished products; inventory costs in GDC	Personnel costs; Equipment costs; floor space costs
	Distribution logistics costs (CD)	Distribution costs	Inventory costs in dealers	Personnel costs; equipment costs; floor space costs

3 Evaluation models

The simulation of the four-echelon supply chain and its results should be carried out based on the most suitable evaluation method. After introducing the zero-based and improved activity-based costing (ABC) approaches, the complexity costs evaluation models corresponding to the KPI are constructed in this section.

3.1 EVALUATION METHODS

A combination of three methods is introduced to evaluate the variant-driven variety complexity costs in automotive supply chains: variety-driven activity-based costing (VD-ABC), as studied by Lechner [1]; time-driven activity-based costing (TD-ABC), as studied by Kaplan and Anderson [29]; and the zero-based approach. The first two methods are both modifications of the ABC approach, while the latter is suitable for dynamic evaluation.

3.1.1 Time-driven activity-based costing and variety driven activity-based costing

Cooper and Kaplan [30] consider that the ABC approach identifies activities in an organization, and assigns the cost of resources for each activity to all products and services according to the actual consumption by each. This method can identify the indirect and direct costs of each activity, as it is helpful to know the profitability of each activity and then reduce or eliminate it in order to optimize overhead costs, as discussed by Goldbach [31].

However, Kaplan and Anderson [29] argue that the ABC approach is not suitable for large-scale organizations as it often fails to capture the complexity of actual operations, takes too long time to implement and is too expensive to build and maintain. In order to resolve these problems, the TD-ABC approach is developed by Kaplan and Anderson [29] by improving and revising ABC.

According to Kaplan and Anderson [29], resource demands are estimated according to the cost per time unit, and only two kinds of parameters are required to be estimated for each group of resources – i.e. the cost per time unit of capacity and the unit times of activities. As opposed to the ABC approach, the TD-ABC approach can help supply chains catch the complexity of business far more simply and find which costs should be saved and which should be cut. It can accommodate the complexity of real operations by time equations, a new feature that enables the model to reflect how order and activity characteristics cause processing times to vary. Time equations greatly simplify the estimating process and produce a far more accurate cost model than traditional ABC techniques, as studied by Kaplan and Anderson [29]. In this section, TD-ABC is used to calculate the variant-driven personnel complexity costs, equipment complexity costs and part of the floor space complexity costs.

However, the TD-ABC approach also has its limitations. Some of the logistics costs cannot be expressed by time equations – for example, the inventory costs and corresponding floor space costs – because the process time of inventory is unrelated to the number of variants. That means it cannot track the influence of

variants on the process time of inventory (see Lechner [1]). Consequently, Lechner [1] expand continually the TD-ABC approach to capture the capacities in parts, and square meter and name it VD-ABC so that it can allocate costs to the level of variants according to the actual input involved. Using the capacity cost rate to derive the costs of variant-driven variety, the VD-ABC can help to determine the number of variety subject to part variants. In this paper, VD-ABC is used to calculate the inventory complexity costs and the floor space complexity costs related to the inventory.

3.1.2 Zero-based approach

The idea of the zero-based approach originates from zero-based budgeting (ZBB), which is an approach to planning and decision-making that works in reverse to that of traditional budgeting. Sarant [32] demonstrates that the zero-based approach requires each budget request to be re-evaluated thoroughly so that each line item of the budget has to be approved, rather than only be changed. The definition of the zero-based approach is grounded in the federal training experience, as shown by Sarant [32], ie that: "A minimum level is actually the grass roots funding level necessary to keep a program alive." As discussed by Sarant [32], the minimal level here is the program or funding level below which it is not feasible to continue a program because no constructive contribution can be made toward fulfilling its objective.

The difficulty of cost management in achieving significant and sustainable cost reductions without threatening performance and competitiveness is discussed by Waterlander et al. [33]. They argue that zero-based cost management can re-examine the activities and associated costs necessary to achieve specific business outcomes so that it can avoid these potential problems. As proposed by Waterlander et al. [33], there are four steps for a successful zero-based approach: re-examining the vision; zero-basing the activities; assessing the outcomes; and embedding the change. The key of the zero-based approach is in justifying the retention or expansion of specific optional activities (and their full cost) for clearly articulated business benefits.

Lechner [1] extend the concept of the zero-based approach as the basic principle underlying variety-driven complexity in automotive inbound logistics. This paper applies the idea of zero-based approach to evaluate the variety complexity costs in whole automotive supply chains, to help them choose the right number of part variants since an affordability line can be used to filter the activities during the operation of the approach. The affordability line is a threshold cost initiated by cost managers to control costs through choosing the suitable number of product variety and part variants. An example of an affordability line is shown in Figure 3.

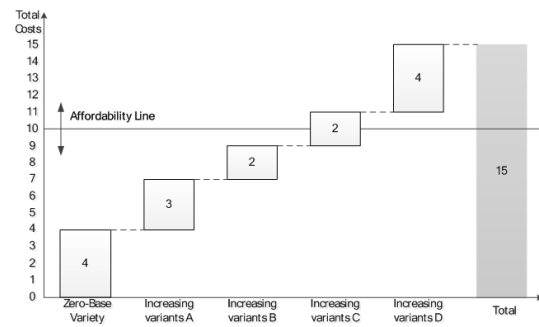


FIGURE 3 Complexity Costs Against a Cost Management Affordability Line

In this paper, a three-scenario model under the supply chain structure in Figure 1 is created, in the light of the zero-based approach, which represents separately that there are no variants, one variant and many variants in part types. The three scenarios are the zero-variant scenario, the one-variant scenario and the rich-variants scenario. The change of complexity costs from the one-variant scenario to the rich-variants scenario against the zero-variant scenario can help a supply chain master the influences of variant and make decisions on the number of variants and variety.

3.1.3 Integrated evaluation steps

Combining the zero-based approach with TD-ABC and VD-ABC, the total complexity costs can be calculated and analyzed. According to the classification of logistics costs in Table 1, the supply chain sets a respective affordability line for each type of costs, as well as for total costs, so that the suitable number of part variants can be decided comprehensively. Depending on the supply chain process and the classification of logistics costs, the integrated evaluation steps can be summed up as follows:

- (1) Calculating every cost item in the three different scenarios: zero-variant, one-variant and rich-variants.
- (2) Calculating logistics/functional costs, logistics processes costs and total logistics costs in the three scenarios.
- (3) Calculating every complexity cost individually in the one-variant scenario and rich-variants scenario, this equals the costs difference between the costs in the latter two scenarios and the zero-variant scenario in (1).
- (4) Calculating logistics/functional complexity costs, logistics processes complexity costs and total logistics complexity costs in the one-variant scenario and rich-variants scenario, which equals the cost difference between the latter two scenarios and the zero-variant scenario in (2).
- (5) Evaluating the complexity costs or choosing the number of part variants according to the affordability lines.

The calculation methods of these costs are introduced in terms of the function in the following analysis since

the same function costs in different processes can be calculated by the same method.

3.2 EVALUATION OF PROCESS COMPLEXITY COSTS

Transport costs and distribution costs can both be regarded as transport costs that are related to transport modes, transport capacity and distance. For the transport of a certain product or part between two specific places, the transport cost CT_{ij} can be expressed by the multiplication of the transport rate r_{ij} and the number of products or parts v_{ij} being transported:

$$CT_{ij} = r_{ij} \cdot v_{ij}. \tag{1}$$

In Equation (1), i refers to the i^{th} product/part, and j refers to the j^{th} route for the transport of the i^{th} product/part. The total transport cost is equal to the sum of all the CT_{ij} :

$$CT_V = \sum CT_{ij}. \tag{2}$$

V in Equation (2) is the total number of part variants. Therefore, the transport complexity cost is:

$$CCT_V = CT_V - CT_0. \tag{3}$$

3.3 EVALUATION OF INVENTORY COMPLEXITY COSTS

The growth of total inventory costs in production/inventory systems is associated with types of supplier–customer systems of supply chains as discussed by Wu [14], and is almost linear in the number of variety as discussed by Benjaafar et al. [17]. One type of supplier–customer system is BTS with lower product variety but high volume, and another is BTO with high variety but low volume, as explored by Wu [14].

The increase of inventory is made up mainly of the stock type and safety stock. In respect to stock type, on the one hand, it is normal that more part variants induce more total volume in stock unless there is an absolute-zero inventory policy. On the other hand, the purchase price of parts becomes higher as the lot sizes decrease. Both these reasons promote the growth of inventory costs.

In respect to safety stock, the safety stock of part variants, work-in-progress (WIP) and finished products are raised to prevent the stock being out. Few safety stocks are needed in the zero-variant scenario because the demand and supply are stable and unchanged, whereas in the rich-variants scenario, due to there being more variants and lower average demands for each variant, the demand volatility is higher, and the plant and dealers have to keep safety stock for each part variant, even product variants. Of course, this implies greater inventory costs.

3.3.1 Calculation of inventory costs

There is a widely accepted formulation for the calculation of inventory costs, i.e. the multiplication of average stock, I_{ik} and unit inventory carrying cost, $r_{ij} \cdot p_i$:

$$CI_{ik} = I_{ik} \cdot r_{ik} \cdot p_i. \tag{4}$$

In Equation (4), i still refers to the i^{th} product/part, k is the location of the warehouse of product/part i , r_{ik} refers to the carrying cost rate of related product/part and p_i is the purchase price of product/part i .

The average stock is related negatively to the turnover rate that is used to measure the quality of the inventory management in the warehouse. It provides information about the capital bound up in stocked products, as shown by VDI [26]. A successful BTO supply chain imposes great requirements on turnover rate in order to respond to customers in time.

3.3.2 Evaluation of variant-driven inventory complexity costs

The inventory equations with v variants are shown in Equation (5), according to the VD-ABC proposed by Lechner [1]:

$$\begin{aligned} I_0 &= \mu_0^0 + z\sigma_0^0 \\ I_1 &= \mu_0^1 + z\sigma_0^1 + \mu_1^1 + z\sigma_1^1 \\ &= I_0 + (\mu_0^1 + z\sigma_0^1 + \mu_1^1 + z\sigma_1^1) - (\mu_0^0 + z\sigma_0^0) \\ &\dots\dots \\ I_V &= \mu_0^V + z\sigma_0^V + \mu_1^V + z\sigma_1^V + \dots + \mu_V^V + z\sigma_V^V = \\ &I_0 + \sum_{v=0}^V (\mu_1^v + z\sigma_1^v) - (\mu_0^0 + z\sigma_0^0). \end{aligned} \tag{5}$$

The parameters in (5) are as follows:

- I_v – inventory in scenario with v variants, $v=0, 1, 2, \dots, V$;
- μ_v^V – mean demand for variant v per unit of time in a scenario with V variants;
- σ_v^V – demand standard variation of variant v per unit of time in a scenario with V variants;
- z – safety parameter chosen from statistics tables to ensure that the probability of this variants’ stockout is above a certain service level.

The inventory cost is a proportion of the value of products in stock, as in Equation (4). The inventory costs with different numbers of variants are shown in Equation (6), as discussed by Lechner [1].

$$\begin{aligned}
 CI_0 &= I_0 \times p_0^0 \times r \\
 CI_1 &= (\mu_0^0 + z\sigma_0^0) \times p_0^0 \times r + [(\mu_0^1 + z\sigma_0^1) \times p_0^1 \times r + \\
 &(\mu_1^1 + z\sigma_1^1) \times p_1^1 \times r - (\mu_0^0 + z\sigma_0^0) \times p_0^0 \times r] = \\
 CI_0 &+ [(\mu_0^1 + z\sigma_0^1) \times p_0^1 \times r + \\
 &(\mu_1^1 + z\sigma_1^1) \times p_1^1 \times r - CI_0] \\
 &\dots\dots \\
 CI_V &= CI_0 + [r \times \sum_{v=0}^V (\mu_v^V + z\sigma_v^V) \times p_v^V - CI_0]
 \end{aligned} \tag{6}$$

CI_v – inventory cost in scenario with v variants, $v=0, 1, 2, \dots, V$;

r – interest rate of holding the stock;

p_v^V – purchase price of variant v , $v=0, 1, 2, \dots, V$.

The simple version of CI_v as in Equation (7) can be obtained from Equation (6):

$$CI_V = r \times \sum_{v=0}^V (\mu_v^V + z\sigma_v^V) \times p_v^V \tag{7}$$

The variant-driven inventory complexity costs with V part variants CCI_V can be calculated from Equation (8), as discussed by Lechner [1], on the premise that the inventory policy will not change with the increase in variants.

$$CCI_V = CI_V - CI_0 \tag{8}$$

The variant-driven inventory complexity costs may occur in the materials inventory during inbound logistics, in the WIP inventory at the original equipment manufacturer (OEM) and in the finished automotive inventory in distribution centres.

3.4 EVALUATION OF RESOURCES COMPLEXITY COSTS

There are not only greater requirements for inventory and transport, but also changes of the process of production and logistics in the total automotive supply chain system due to the proliferation of variant-driven variety. Some additional processes are added to the existing production process in order to mitigate the mistakes or reworks when the variants increase to a certain level; and a more complicated production process may cause a reduced level of automation in return. Both these impacts prolong the whole production process. Meanwhile, in order to solve the problem that floor space beside an assembly line is not sufficient to deal with too many variants, some additional processes have to be added before assembly outside the assembly shop. The increase of processes

brings demands for additional personnel, space and equipment resources. The costs of personnel, floor space and equipment are all related to the process time, and so can be evaluated by TD-ABC. The summation of the three kinds of costs with V part variants is regarded as the resources costs, CR_v . Accordingly, the resources complexity costs are CCR_v .

3.4.1 Production process time

Since there are some relation formulae between process time and resources costs, process time and incremental process time should be calculated first. The process time equation with different numbers of variants in this project can be shown as Equation (9), according to TD-ABC as discussed by Lechner [1].

$$\begin{aligned}
 TP_0 &= t_0 x_0^0 \\
 TP_1 &= t_0 x_0^1 + t_1 x_1^1 = t_0 x_0^0 + (t_0 x_0^1 + t_1 x_1^1 - t_0 x_0^0) \\
 &\dots\dots \\
 TP_V &= t_0 x_0^V + t_1 x_1^V + \dots + t_V x_V^V \\
 &= t_0 x_0^0 + \sum_{v=0}^V t_v x_v^V - t_0 x_0^0
 \end{aligned} \tag{9}$$

t_0 – process time in the zero-variant scenario;

t_v – incremental process time of incremental activity v , $v=1, 2, \dots, V$;

TP_v – time for process in a logistics system with v variants;

x_v^V – parts number of variant v in products with V variants.

This equation makes sure that the process time is allotted to individual variants. The fixed-step cost characteristics can also be shown by this equation. These occur if there is no margin for existing capacities, and new processes have to be added to expand the capacities.

Set ΔTP_v is the incremental process time between the scenario with V variants and the zero-variant scenario, as in Equation (10):

$$\Delta TP_V = TP_V - TP_0 \tag{10}$$

3.4.2 Evaluation of personnel and equipment complexity costs

The personnel complexity costs CCP_V and equipment complexity costs CCE_V can be calculated through multiplying the process time by correspondent cost rates according to TD-ABC, as proposed by Kaplan and Anderson [29]. The equipment here includes both equipment and auxiliaries.

$$CCP_V = \Delta TP_V \cdot CRP, \tag{11}$$

$$CCE_V = \Delta TP_V \cdot CRE. \tag{12}$$

CRP – cost rate of the capacity supply of personnel;
CRE – cost rate of capacity supply of equipment.

The key point here is to count completely the increasing process time in every segment. These times may include transport and delivery, loading and unloading, sequencing, unpacking and repacking, etc.

3.4.3 Evaluation of floor space complexity costs

There are two kinds of requirements for more floor space with an increase of part variants in automotive supply chains. The first, $CCF_{p,v}$, is caused by the increase of processes for supporting the production process. The second, $CCF_{s,v}$, derives from keeping the incremental stocks of part variants and finished products. These will be examined in turn.

First, for floor space complexity costs for additional processes, Lechner [1] analyzes the interaction between total time consumed and floor space in different working processes. She certifies that the improvement in processes, i.e. from standard delivery to delivery with internal preparation to just in sequence (JIS) delivery, can change the requirements for floor space beside the assembly line through extending the whole process time.

There should be floor space for the implementation of new increased processes. These floor spaces, which may be located in the internal warehouse, beside the assembly line or in the external warehouse, are closely related to the containers and equipment used during such processes. Hence, the container areas can be used to estimate the floor space needed by new processes:

$$F_{p,1} = r \left(c_1^V \times f_1^V \right) \left(w_1^V \times l_1^V \right) \tag{13}$$

$$F_{p,V} = r \sum_{v=1}^V \left[\left(c_v^V \times f_v^V \right) \left(w_v^V \times l_v^V \right) \right]$$

$F_{p,v}$ – area of floor space induced by additional processes with v variants, compared with the zero-variant scenario, $v=0, 1, 2, \dots, V$;

w_v^V – width of the containers for variant v ;

l_v^V – length of the containers for variant v ;

r – magnification coefficient of floor space when considering the demands of activity space for equipment and workers;

c_v^V – number of containers for variant v ;

f_v^V – binary variable to express whether an incremental process is required or not with v variants. If it is required, $f_v^V = 1$; otherwise, $f_v^V = 0$.

Then, the first kind of floor space complexity costs can be calculated in Equation (14), as shown by Lechner [1]:

$$CCF_{p,V} = F_{p,V} \cdot CRF_p. \tag{14}$$

$CRF_{p,v}$ – floor space cost rate for the additional process area.

Second, for floor space complexity costs for increased stocks, it is difficult to estimate these costs because there are several possible solutions for a company to obtain the floor space, such as through renting, construction or third-party logistics (TPL) management. It should be assumed that all factories rent warehousing to satisfy their demands for floor space in this project. The floor space complexity costs can be calculated through multiplying the lease costs and space demands. As for floor space, it can be estimated by stock level, container/pallet size, warehouse height and the effective area utilization factor:

$$F_{s,V} = \sum_{v=0}^V \frac{I_{v,\max}}{f_V L_V} \cdot (w_v \cdot l_v). \tag{15}$$

$F_{s,v}$ – floor space demands for stocking area of V variants, $v=0, 1, 2, \dots, V$;

f_v – effective area utilization factor of warehouse, which refers to the ratio of storage area to warehouse area;

$I_{v,\max}$ – maximum stock of variant v in warehouse;

L_v – stack layers of container/pallet of variant v in warehouse.

$L_v = [H / h_v]$. $[H / h_v]$ is the maximum integer that is less than or equal to H / h_v . H refers to the height of warehouse. The parameters, w_v , l_v and h_v are the width, length and height of the container/pallet of variant v , respectively.

The floor space costs $CF_{s,v}$ and the complexity costs of floor space $CCF_{s,v}$ can be achieved through Equation (16) and (17):

$$CF_{s,V} = CRF_s \cdot F_{s,V}, \tag{16}$$

$$CCF_{s,V} = CF_{s,V} - CF_{s,0}. \tag{17}$$

CRF_s – monthly floor space cost rate for the incremental area.

Just as with the inventory complexity costs, the floor space complexity costs may also occur in the materials inventory during inbound logistics, WIP inventory in OEM and the finished automotive inventory in distribution centres.

3.5 TOTAL COMPLEXITY COSTS

Obviously, the total logistics costs CV and complexity costs CCV are, above all, the sums of the corresponding costs, which are shown in Equations (18) and (19):

$$CV = CT_V + CI_V + CR_V = CT_V + CI_V + \left(CR_V + CE_V + CF_{p,V} + CF_{s,V} \right), \quad (18)$$

$$CCV = CCT_V + CCI_V + CCR_V = CCT_V + CCI_V + \left(CCP_V + CCE_V + CCF_{p,V} + CCF_{s,V} \right), \quad (19)$$

where CP_V and CE_V are the personnel costs and equipment costs with V variants individually.

4 Simulation with OTD-NET and analysis of the results

In this section, the four-echelon automotive supply chain model in Figure 1 is created firstly by the OTD-NET. Three different scenarios are then simulated separately, and the complexity costs are calculated and analyzed in order to evaluate the impacts of variants on the supply chain.

4.1 SCENARIO-BASED PROCESS MODEL

The discrete event simulation tool OTD-NET was developed by the Fraunhofer Institute for Material Flow and Logistics (IML). OTD-NET has been applied successfully in many logistics projects (e.g. Intelligent Logistics for Innovative Product Technologies, ILIPT, for VW commercial vehicles) and in the dissertations of Cirullies [34] and Klingebiel [35]. It is especially suited for use in the field where information flow is tightly coupled with material flow, as shown by Cirullies [34] and Klingebiel [35]. The initial aim of OTD-NET was to serve the automotive industry, and it has now gained experience in application and improvement in this industry over many years. It was able to simulate the diversity and complexity in the automotive supply chain necessary for this paper. Therefore, we chose OTD-NET as the simulation tool to obtain our evaluation results. The modelling structure of the scenario supply chain is shown in Figure 4.

As an object-oriented simulation tool, OTD-NET contains the classes “supplier”, “plant” and “dealer”. Hence, these supply chain nodes can be created directly in the simulation model, as demonstrated in Figure 4. However, there are no classes such as “warehouse” and “distribution center” in OTD-NET. It provides other innovative concepts: “buffer” (including “inbuffer”, “outbuffer” and “goods receipt buffer”), “routing” and

“distribution channel”. “Buffer” (expressed by the green chevron-shaped symbol in Figure 4) and “distribution channel” (expressed by the blue chevron-shaped symbol) are both used to emphasize that goods should be always in the flow process but not in stagnancy before they reach the ultimate customers, as found by Wagenitz [36] and Li [37]. The warehouses are represented by “inbuffer” and “outbuffer”, respectively, according to their functions and positions in this project. “Inbuffer”, in which the containers of inbound logistics are provided, must be created when modeling an inbound logistics system. Wagenitz [36] proposes that “outbuffer” is defined as the starting point of the inbound logistics between the second- and first-tier suppliers, or between the supplier and the plant. The GDC is also represented by “buffer” in this project.

“Distribution channel” is one of the most important classes used to model the distribution, as discussed by Wagenitz [38]. Wagenitz [36] and Li [37] argue that a “distribution channel” can represent an actual route channel from starting point to end point located in different positions (e.g. from suppliers to plant, from plant to dealers), or the route in a logistics node such as in a warehouse or distribution center. It describes the transport time and transport capacities between the locations as discussed by Wagenitz [36]. However, Wagenitz [36] also claims that if there are several alternative routes between logistics nodes in two echelons, a class “routing table” (expressed with a diamond-shaped symbol in Figure 4) and its attribute “routing” (expressed with grey chevron-shaped symbol) will be used to choose an alternative distribution channel (e.g. air cargo instead of seaborne transport). Each “route” has a corresponding “distribution channel”.

From Figure 4, it can be shown that the group body in white comes into being after production in the welding shop. Therefore, two layers of suppliers are built here. The combination of press shop and welding shop is regarded as a BTO supplier (i.e. body shop) and the body sheets supplier is the BTS supplier of the body shop. In order to simplify the model, the paint shop and assembly shop are combined into one plant: “paint and assembly plant” in the model.

4.2 EVALUATION OF COMPLEXITY COSTS OF VARIANT-DRIVEN VARIETY

In the three scenarios of zero-variant, one-variant and rich-variants, in the zero-variant scenario there is only one kind of base part in each part type, which means only one kind of base product is produced. This product has a 100% actual rate in the market, and the actual rates of all the parts are also 100%. In the one-variant scenario, one part variant is added in the part type group paint. The actual rates of the two kinds of paint are 0.6 and 0.4. Therefore, two kinds of automotive are produced in this scenario. In the rich-variants scenario, there are a couple of part variants in each part type. In detail, there are 11

kinds of paint, six kinds of engine, four kinds of transmission, seven kinds of seat, two kinds of door and locks, three kinds of wheels and tires, and three kinds of body in white as the unfinished products in the body shop. All these part variants have different actual rates.

Therefore, the total number of product variants is equal to the arithmetic product of all the number of part variants, that is $3 \times 11 \times 6 \times 4 \times 7 \times 2 \times 3 = 33,264$ if all the part variants in one part type can match with all the part variants in other part types.

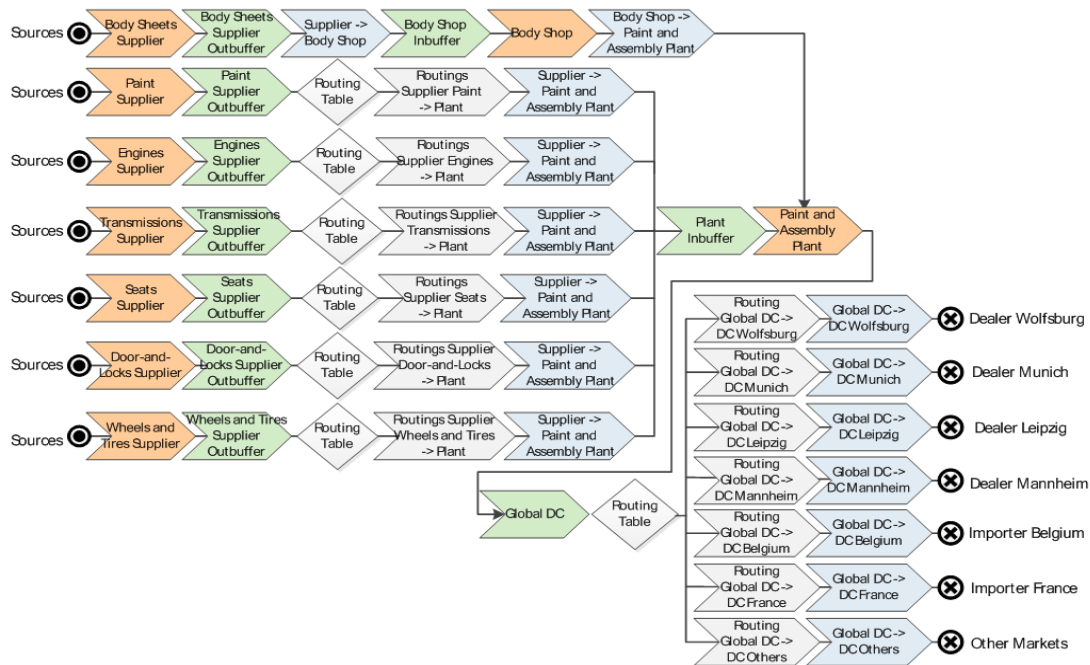


FIGURE 4 OTD-NET modelling structure of the scenario supply chain

The production cycle time of a car is 2,340 minutes in the zero-variant and one-variant scenarios, but it goes up to 2,640 minutes in the rich-variants scenario. Assume the market demands for a car in one year shown in Table 2. By running the simulator, all the complexity costs can be calculated with the evaluation models in the previous section.

TABLE 2 Total market demand for PC car

Month	1	2	3	4	5	6
Demand	18000	13500	13500	15000	18000	15000
Month	7	8	9	10	11	12
Demand	15000	12000	10500	9000	12000	13500

TABLE 3 Total variant-driven variety costs

	CT	CI	CR			CV	
			CP	CE	CF _p		CF _s
Scenario 1	40472706	12151358	134672265	38477790	0	363038	226137157
Scenario 2	40483628	12030035	134721405	38491830	0	359790	226086688
Scenario 3	40275500	22259627	151110036	43174296	1866.24	570701	257392026

TABLE 4 Total variant-driven variety complexity costs

	CCT	CCI	CCR			CCV	
			CCP	CCE	CCF _p		CCF _s
Scenario 2	10922	-121323	49140	14040	0	-3248	-50468,51
Scenario 3	-197206	10108269	16388631	4682466	1866,24	207663	31191689

The simulation results of all the complexity costs can be summarized as follows, according to the calculation results.

(1) Process complexity costs. The number of parts and finished cars transported in the three scenarios is almost

unchanged because the productions are similar. Therefore, the process costs are approximately the same. The process costs change very slightly, meaning the impact of variants on process costs is very small.

The impacts of increasing variants on transport costs are so insignificant as to be negligible if the total market demands or outputs remain unchanged. The total demands for parts and products will not change either under such a situation, so the transport burden on parts and products is the same. What will be altered are the transport frequency and transport volume of each part variant at each time.

(2) Inventory complexity costs. The stocks of body in white and unfinished cars are not considered because normally they are kept on the production lines and do not produce demands for additional resources. Only the inventory complexity costs of parts and finished cars are analyzed in this paper. The number of part variants of body sheets is the same in the three scenarios, so the inventories are all the same. All the other numbers of part variants in scenario 2 is equal to that in scenario 1, except for paint. That is why the average inventories in scenario 2 are comparatively close to that in scenario 1. In scenario 3, as compared to the other two scenarios, the number of every part variant increases greatly, apart from body sheets, which leads to a sharp rise in the inventory. Completely different process costs and impact of variants on inventory costs are manifest. The inventory costs of parts change slowly when there are few variants, while they increase quickly with greater variety due to the multiplicative effect of more part variants.

The characteristics of the BTO supply chain maintain the stock of cars at a low level. Moreover, all the variants of the finished automobiles ordered by one dealer are distributed through one distribution channel, which are distinctive to the transportation of parts: each part variant is transported through a special distribution channel. The gaps of stocks of finished cars among the three scenarios are not as huge as those of parts. Therefore, there is not so great an impact of variant-driven variety on inventory costs of finished cars as on part inventory costs. In particular, when there are only few variants, the inventory costs of automobiles change very little (and sometimes decrease slightly).

(3) Resources complexity costs. Personnel and equipment complexity costs are both related directly to the total process time. Although the values of the two kinds of costs above are completely different, their variation tendencies are the same. That is, an increase in variants has the same impact on personnel costs and on equipment costs. In the model constructed in this project, an additional process, sequencing, occurs in scenario 3 compared with scenarios 1 and 2. That means the requirements for floor space on processing in scenario 2 are the same in scenario 1. All the other five-part types, except for body sheets and paint, should be sequenced before being fed to the assembly line. Therefore, floor space complexity costs for additional processes only occur in scenario 3. As for floor space complexity costs for increased stocks, they are decided by the maximum stock of parts and finished cars. The variation of floor space costs for storing parts conforms to the change of

maximum stocks. Similar to the change of inventory, the maximum stocks in scenario 2 are very close to those in scenario 1, while the maximum stocks in scenario 3 increase significantly. Obviously, that phenomenon is the result of the huge increment of variants from scenario 2 to scenario 3. However, the impacts of variants on maximum stocks of cars are not as substantial as on part stocks.

Furthermore, because of the decrease of production volume and the difference of transportation volume, several complexity costs in Table 4 are negative. Nevertheless, it can be shown from Table 4 that the total costs increase greatly. Table 5 shows the growth rate of each complexity cost. According to Tables 4 and 5, the total logistics costs in scenario 3 increases by about 13.8% from that of scenario 1. Figure 5 shows the costs in Table 3 using a bar chart. It is obvious from Figure 5 that resources costs account for the majority of total logistics costs compared with the other two kinds of costs, while it can also be seen that inventory complexity costs in scenario 3 account for about 83.186% of the inventory costs in scenario 1 – i.e. the growth rate of inventory costs is 83.186% between scenario 1 and scenario 3. Inventory costs change slowly when there are few variants, while they ascend quickly with the propagation of variety because of the multiplicative effect of more part variants. In addition, the impact of parts on inventory costs is greater than that of finished cars.

TABLE 5 Growth rates of logistics costs compared with scenario 1

	<i>CCT</i>	<i>CCI</i>	<i>CCR</i>	<i>CCV</i>
Scenario 2	0.027%	-0.998%	0.035%	-0.022%
Scenario 3	-0.487%	83.186%	12.265%	13.793%

To sum up, the proliferation of variant-driven variety had a huge influence on the increase of logistics costs, especially on inventory costs and resources costs. Hence, managers should pay more attention to the management of inventory costs and resources costs during an increase in the number of variants.

5 Conclusion

With the objective of helping automotive manufacturers optimize the number of part variants and product variety in respect of saving costs, this project studies the methods for evaluating the complexity costs caused by the proliferation of part variants on automotive supply chain. First, a four-echelon automotive supply chain model from suppliers to dealers is constructed with a process chain model, and the KPI system of evaluating complexity costs is achieved. Second, combining the TD-ABC, VD-ABC and zero-based approaches, the evaluation models of all complexity costs in regard to KPI are constructed. These complexity costs are process complexity costs, inventory complexity costs and resources complexity costs. Lastly, the four-echelon supply chain model is created with an OTD-NET simulator and three scenarios

with different number of variants are simulated. The results of these three scenarios are used to evaluate the impacts of variants on the automotive supply chain dynamically.

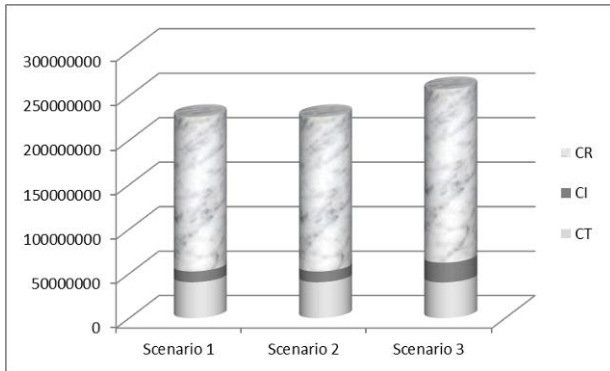


FIGURE 5 Components of total variant-driven variety costs

The research shows that the proliferation of part variants has a great impact on the logistics costs of automotive supply chains. Inventory complexity costs are particularly influenced by the variant-driven variety. Inventory costs and resources costs are also influenced significantly. In the light of the zero-based approach, the automotive manufacturing can choose the suitable number of variants by referring to the affordability line. The results of this project are also helpful for other industries to evaluate problems of variety.

References

- [1] Lechner A 2011 *Modellbasierter Ansatz zur Bewertung vielfaltsinduzierter Logistikkomplexität in der variantenreichen Serienfertigung der Automobilindustrie* Unternehmenslogistik. Verlag Praxiweissen: Dortmund, Germany
- [2] Yi H Y, Klingebiel K 2013 Contribution to an automotive variant-driven product variety framework: influencing factors and impacts *Proceedings of the Production and Operational Management Conference*
- [3] Klingebiel K 2008 Framework of OPP influencing factors *Working paper in Technology University Dortmund*
- [4] Ramdas K 2003 *Production and Operations Management*, 12(1) 79-101
- [5] Fisher M, Jain A, MacDuffie J P 1995 Strategies for product variety: lessons from the auto industry. Kogut B, Bowman E Eds. *Redesigning the Firm* Oxford University Press: New York, USA, 116-54
- [6] Schleich H, Lindemann E, Miemczyk J, Stone G, Holweg M, Klingebiel K, Seewald J, Esterházy M, Kauth, J, Seidel T 2005 D7/8.1.3 Analysis of state of the art of complexity management *Working paper (1999-12259) ILIPT Project*: Nuremberg, Germany
- [7] Fisher M L, Ittner C D 1996 *The Impact of Product Variety on Automobile Assembly Operations: Analysis and Evidence*: Knowledge@Wharton April 15
- [8] Fisher M L, Ittner C D 1999 *Management Science* 45(6) 771-85
- [9] Ittner C D, Macduffie J P 1995 *Production and Operations Management* 4(4) 312-86
- [10] Jagersma P K 2008 *Business Strategy Series* 9(5) 238-42
- [11] Thonemann U W, Bradley J R 2002 *European Journal of Operational Research* 143(3) 548-69
- [12] MacDuffie J P, Sethuraman K, Fisher M L 1996 *Management Science* 42(3) 350-69
- [13] Scavarda L F, Reichhart A, Hamacher S, Holweg M 2010 *International Journal of Operations & Production Management* 30(2) 205-24
- [14] Wu L F 2007 *International Journal of Management* 24(3) 499-509
- [15] Kersten W, Rall K, Meyer C M, Dalhöfer J 2006 Complexity management in logistics and ETO supply chains Blecker T, Kersten W *Complexity Management in Supply Chains: Concepts, Tools and Methods* Erich Schmidt Verlag GmbH & Co: Berlin, Germany, 325-42
- [16] Roy R, Evans, R, Low, M J, Williams D K 2011 *Journal of Engineering Manufacture* 225(10) 1939-50
- [17] Benjaafar S, Kim J S, Vishwanadham N 2004 *Annals of Operations Research* 126(1-4) 71-101
- [18] Fujimoto H, Ahmed A, Iida Y, Hanai M 2003 *International Journal of Flexible Manufacturing Systems* 15(4) 283-307
- [19] Perona M, Miragliotta G 2004 *International Journal of Production Economics* 90(1) 103-15
- [20] Pil F K, Holweg M 2004 *Interfaces* 34(5) 394-403
- [21] Schoeller N 2007 *Translation of Complexity Management Journal* (2) 13-15
- [22] Berry W L, Cooper M C 1999 *Journal of Operations Management* 17(2) 163-78
- [23] Scavarda L F, Schaffer J, Scavarda A J, Reis AC, Schleich H 2009 *Benchmarking: An International Journal* 16(3) 387-400
- [24] Kuhn A 1995 *Prozessketten in der Logistik: Entwicklungstrends und Umsetzungsstrategien*. Unternehmenslogistik. Verlag Praxiweissen: Dortmund, Germany
- [25] Kuhn A, Hellgrath H, 2002 *Supply Chain Management: Optimierte Zusammenarbeit in der Wertschöpfungskette*, Springer: Berlin and Heidelberg, Germany
- [26] Verein Deutscher Ingenieure 2002 *Logistikkennzahlen für die Distribution* Blatt 3 Beuth Verlag GmbH, VDI 4400: Düsseldorf, Germany
- [27] Karmarkar U S 1996 *Journal of Marketing Research* 33(2) 125-33
- [28] Shapiro B P 1977 *Harvard Business Review* 55(5) 104-14

- [29]Kaplan R S, Anderson S R 2004 *Harvard Business Review* November 1-8
- [30]Cooper R, Kaplan R S 1988 *Harvard Business Review* September/October 96-103
- [31]Goldbach M 2002 Organizational settings in supply chain costing Seuring S, Goldbach M Eds *Cost Management in Supply Chains* Physica-Verlag: Heidelberg 89-108
- [32]Sarant P C 1978 *Zero-Base Budgeting in the Public Sector: A Pragmatic Approach*, Addison-Wesley Publishing Company: Boston USA
- [33]Waterlander O, Clark A, Groves S, Nordahl H 2010 *Zero-based cost management: a holistic approach to managing budgets* Booz & Company
- [34]Cirullies J 2009 *Erprobung and Bewertung eines simulationsbasierten Ansatzes zur Bewertung langfristiger Gestaltungsmaßnahmen hinsichtlich der Wirkung auf die Robustheit von Distributionsnetzwerk* Diplomarbeit, TU-Dortmund: Dortmund, Germany
- [35]Klingebliel K 2009 *Entwurf eines Referenzmodells für Built-to-order-konzepte in Logistiknetzwerken der Automobilindustrie. Unternehmenslogistik* Verlag Praxiswissen: Dortmund, Germany
- [36]Wagenitz A 2007 *Medellierungsmethode zur Auftragsabwicklung in der Automobilindustrie. Unternehmenslogistik.* Verlag Praxiswissen: Dortmund, Germany
- [37]Li C 2012 *An integrated Approach to robust Multi-Echelon Inventory Policies for Distribution Networks* Unternehmenslogistik Verlag Praxiswissen: Dortmund, Germany

Authors	
	<p>Haiyan Yi, born in December, 1978, Hunan Province, P.R. China</p> <p>Current position, grades: Associate Professor of School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China. University studies: PhD from Southwest Jiaotong University, China. Scientific interest: supply chain risk management, logistics planning. Publications: more than 20 papers published in various journals. Experience: research and consulting experiences of 7 years, 3 scientific research projects.</p>
	<p>Dianjun Fang, born in October, 1961, Shandong Province, P.R. China</p> <p>Current position, grades: Professor of Chinese-German School for Postgraduate Studies, Tongji University, Shanghai, China. University studies: PhD from the University Dortmund, Germany. Scientific interest: logistics and supply chain management. Publications: more than 30 papers published in various journals. Experience: teaching experience above 14 years, research, consulting and training experience above 20 years.</p>

Optimal acquisition and pricing policies for remanufacturing systems with initial investment

Chenghu Yang¹, Lanying Liu¹, Lei Zhang^{1, 2*}

¹All-Trans Logistics School, Fuzhou University, Fuzhou, 350108, China

²College of Economics and Management, Guangxi Teachers Education University, Nanning 530001, China

Received 1 March 2014, www.tsi.lv

Abstract

The problem of used product (core) acquisition is an important issue in remanufacturing. In traditional models, remanufacturing systems are assumed to be well-established. However, remanufacturing systems in most developing countries are imperfect and remanufacturers need to make huge initial investments to improve the remanufacturing systems. It is therefore suggested that the effects of initial investment cannot be neglected. In this paper, an acquisition and pricing problem in imperfect remanufacturing systems is studied. The problem is firstly formulated as a two-period nonlinear programming model, and the closed forms of the optimal solution are presented based on Karush-Kuhn-Tucker conditions. Next, the multi-period acquisition and pricing problem, and the effects of the initial investment are discussed. Finally, the conclusions are testified by numerical examples. The results show that with the remanufacturing system improved, the remanufacturer will increase investment and acquire more cores.

Keywords: Core acquisition, Imperfect remanufacturing systems, Nonlinear programming model, Karush-Kuhn-Tucker conditions

1 Introduction

Remanufacturing is growing into a booming industry because of its resource sustainability and environmental friendliness. Remanufacturing driven by cost reduction has become an important technological field and claims a large market share, especially in developed countries. However, in developing countries (e.g., Brazil, China, and India), there are different levels of development in remanufacturing. Remanufacturing systems in developing countries suffer from limitations, such as a poor recycling network, deficiency in regulation, low customer recognition and low environmental awareness. These limitations reduce the profitability and enthusiasm of firms. In order to improve remanufacturing systems, the initial investment should be made based on the recovery network, remanufacturing technique, recovery advertising, and so on. The well-established remanufacturing system depends on huge initial investment and long-term development, and consequently most firms undertake their remanufacturing activities on the grounds of environmental responsibilities rather than economic benefits.

Motivated by the above observations, an acquisition and pricing problem is studied in the presence of initial investment. In order to determine the optimal policies, the two-period and multi-period models are developed respectively. Moreover, the effects of the initial investment on the optimal policies are analysed. Finally, numerical experiments are provided to validate the analysis conclusions.

The rest of this paper is organized as follows: In section 2, related research is briefly reviewed. In section 3, the problem is described and a two-period model is formulated. Additionally, the effects on the optimal policies are analysed. In section 4, the model and the conclusions are extended to a multi-period model. In addition, the numerical examples are illustrated in section 5. Finally, section 6 concludes the paper with future research directions.

2 Literature review

There is an abundance of literature related to remanufacturing, which can be categorized into different research themes including inventory management [1, 2], production planning and control (PP&C) [3], capacity planning [4, 5], product acquisition [6, 7], and network design [8, 9].

The literature related to our work is reviewed. Debo et al., Heese et al., and Ferguson and Toktay studied the policies as to whether to manufacture new products and remanufactured products or not [10, 12]. Ferrer and Swaminathan developed two-period, multi-period, and infinite planning models respectively to identify thresholds in remanufacturing operations in the monopoly and duopoly environment [13]. These literature assumed that the new and remanufactured products were undistinguishable. Ferrer and Swaminathan further extended the model to the case in which consumers could differentiate between remanufactured products and new products [14]. Webster and Mitra studied a general two-

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

period model to investigate the effects of take-back laws on remanufacturing [15]. Additionally, Mitra and Webster continued their work to focus on the impacts of government subsidies [16].

In the existing literature, it is assumed that the remanufacturing system is well-established with a perfect recovery network and that consumers have a high awareness of recycling. Apparently, these assumptions are in conflict with the real situation, which is that firms have high incentive to remanufacture cores, especially in developing countries. In these developing countries, due to the imperfect remanufacturing system, the impact of initial investment cannot be neglected. In addition, the existing literature assumes that the return rate is exogenous and is generally set to the statutory return rate. However, in the imperfect remanufacturing system, the statutory return rate could not be set or strictly enforced. Firms could determine the quantity of the remanufactured products, and the return rate is no longer exogenous. In this paper, the initial investment is incorporated into a pricing problem for a remanufacturing system. Most importantly, it is focused on the impacts of the initial investment on optimal policies.

3 The two-period acquisition and pricing model

3.1 PROBLEM DESCRIPTIONS AND ASSUMPTIONS

This paper assumes that the analysis is applied in an imperfect remanufacturing system. In this remanufacturing system, a monopolist firm manufactures new products in the first period and makes both new products and remanufactured products in the future period. For the multi-period case, there are M planning horizons. At the beginning of the first period, the firm determines the optimal acquisition and pricing of each period. The analysis is based on the following assumptions:

Assumption 1. There is no distinction between new products and remanufactured products. That is, consumers cannot distinguish remanufactured products from new products, and the firm charges the same prices for both products.

Assumption 2. Q is the size of the potential market, which is constant in each period. c is the unit cost of new products. s is the unit cost saving of remanufactured products, so the unit cost of remanufactured products is $c - s$.

Assumption 3. p_{ij} is the unit price of j type of product in period i ($i = 1, 2, \dots, M$), and q_{ij} is the quantity of the corresponding product type. Subscript j will take the values N, R , and A , denoting new products, remanufactured products and all products, respectively.

Note that the subscript j is omitted in the first period with only new products.

Assumption 4. In period i , remanufactured products are made of cores sold in period $i-1$, and will be remanufactured and sold in the present period.

Let p_i ($i = 2, \dots, M$) denote the core return rate in period i , $0 < \rho_i < 1$. Therefore, $q_{2R} \leq \rho_2 q_1$ in the two-period case and $q_{iR} \leq \rho_i(Q - p_{i-1})$ in the multi-period case.

Since the remanufacturing system is deficient, the firm should invest in order to improve the system. Firms have little incentive to remanufacture products. The return rate is endogenous variable. Furthermore, the less effective the system, the higher the investment for the same return rate. Hence, the return rate function proposed by Savaskan et al. is extended [17].

Assumption 5. The function between the investment and the return rate is $I_i = (K_i / 2) \rho_i^2$ in period i ($i = 2, \dots, M$). K_i is a scaling parameter, which represents the investment per return rate. It is obvious that the scale parameter decreases with the improvement of the remanufacturing system. That is, $K_{i-1} > K_i$.

Assumption 6. The demand function is linearly decreasing in price. Hence, the demand for period i is $q_{iA} = q_{iN} + q_{iR}$. Additionally, β denotes the discount factor per period, and $0 < \beta < 1$.

To avoid trivial cases, it is assumed that $Q > c$, $c > s$, $2K_i > \beta s^2$.

3.2 THE MODEL FORMULATION AND PROPERTIES

In the two-period case, the firm manufactures new products in the first period and manufactures both new and remanufactured products in the second period. The quantity of new products in the first period is given by $q_1 = Q - p_1$, and the total quantity of new products and remanufactured products in the second period is given by $q_{2A} = q_{2N} + q_{2R} = Q - p_2$. Hence, considering the investment in the remanufacturing system, the model can be expressed as:

(Model P1)

$$\max_{p_1, p_2, q_{2N}, q_{2R}} \Pi_{2R} = (p_1 - c)q_1 + \beta\{(p_2 - c)q_{2N} + (p_2 - c + s)q_{2R} - K\rho^2 / 2\}, \tag{1}$$

s.t.:

$$q_{2R} \leq \rho_2 q_1 \tag{2}$$

In P1, $(p_1 - c)q_1$ and $(p_2 - c)q_{2N}$ denote the profits on new products in the first period and the second period respectively. $(p_2 - c + s)q_{2R}$ is the profit on remanufactured products in the second period, and the subscript in this term ($i = 2$) is omitted.

Let λ be the according dual variable of constraint (2). The Karush-Kuhn-Tucker (KKT) conditions of the acquisition problem can then be rewritten as follows:

$$(Q - p_1) - (p_1 - c) - \lambda\rho = 0, \tag{3}$$

$$\beta s - \lambda = 0, \tag{4}$$

$$-\beta k\rho + \lambda(Q - p_1) = 0, \tag{5}$$

$$\lambda[\rho(Q - p_1) - q_{2R}] = 0. \tag{6}$$

Thus, the optimal solution is shown in Table 1.

TABLE 1 The optimal solution of P1

Policies	Period 1	Period 2
Pricing	$p_1 = \frac{Q+c-\beta s\rho}{2}$	$p_2 = \frac{Q+c}{2}$ $q_2 = \frac{Q-c}{2}$
Quantity	$q_1 = \frac{Q-c+\beta s\rho}{2}$	$q_{2N} = \frac{(Q-c)(1-\rho)-\beta s\rho^2}{2}$ $q_{2R} = \frac{(Q-c)\rho+\beta s\rho^2}{2}$
Acquisition		$\rho^* = \frac{s(Q-c)}{2K-\beta s^2}$

Due to $q_{2N} > 0$, $(2K - \beta s^2)^2 > 2Ks(Q - c)$. Next, the effects of initial investment on the optimal solution are analysed.

Proposition 1. (a) The optimal return rate is the decreasing function of K . (b) I is the decreasing function of K .

Proof. (a) As shown in the expressions of ρ^* , the first order condition is:

$$\left. \frac{\partial \rho}{\partial K} \right|_{\rho=\rho^*} = \left. \frac{-2s(Q-c)}{(2K-\beta s^2)^2} \right|_{\rho=\rho^*} < 0.$$

So, ρ^* is the decreasing function of K .

(b) Due to $I = (K/2)\rho^2 = \frac{K}{2} \left[\frac{s(Q-c)}{2K-\beta s^2} \right]^2$, it is easily

concluded that the first order condition $\frac{\partial I}{\partial K} = \left[\frac{s(Q-c)}{2K-\beta s^2} \right]^2 \left[\frac{-2K-\beta s^2}{2K-\beta s^2} \right] < 0$. Therefore, I is the decreasing function of K .

Proposition 1 implies that the lower the scaling parameter, the higher the return rate. Similarly, the lower the scaling parameter is, the higher the total investment. The lower scaling parameter means that the remanufacturing system is well-established. That is, the investment for the same return rate is relatively small. In the market-driven remanufacturing system, the firm has higher incentive to acquire and remanufacture cores. Additionally, the well-established remanufacturing system promotes consumers' awareness of remanufacturing. Hence, the firm increases the

investment on remanufacturing activities and network. This conclusion coheres with the real situation. In developed countries, not only the related laws and regulations, but also the recovery network and consumer awareness are well-established, which leads to a higher optimal return rate, for example, some firms in America and Germany can reach a rate of 90%. While the remanufacturing industry in developing countries is still in the difficult beginning period, the real situation, such as the recovery network, firms' enthusiasm, society consumer awareness, and so on, need to be improved. The return rate is therefore small and it is hard to obtain investment.

Proposition 2. (a) p_1 is the increasing function of K . (b) q_1 is the decreasing function of K . (c) The investment has no impact on p_2 and q_2 . (d) q_{2N} is the increasing function of K , and q_{2R} is the decreasing function of K .

Proof. (a) The first order condition of p_1 could be written as follows:

$$\frac{\partial p_1}{\partial K} = \frac{\partial p_1}{\partial \rho} \cdot \frac{\partial \rho}{\partial K} = \frac{-\beta s}{2} \cdot \frac{-2s(Q-c)}{(2K-\beta s^2)^2} = \frac{-\beta(Q-c)s^2}{(2K-\beta s^2)^2}.$$

Since $\beta > 0$ and $Q - c > 0$, $\frac{\partial p_1}{\partial K} > 0$. That is, p_1 is the increasing function of K .

(b) Similarly,

$$\frac{\partial q_1}{\partial K} = \frac{\partial q_1}{\partial \rho} \cdot \frac{\partial \rho}{\partial K} = \frac{\beta s}{2} \cdot \frac{-2s(Q-c)}{(2K-\beta s^2)^2} = -\frac{\beta s^2(Q-c)}{(2K-\beta s^2)^2} < 0.$$

Hence, q_1 is the decreasing function of K .

(c) As shown in the expressions of p_2 and q_2 in table 1, the change is not dependent on K .

(d) According to the first order condition, we have:

$$\frac{\partial q_{2N}}{\partial K} = \frac{\partial q_{2N}}{\partial \rho} \cdot \frac{\partial \rho}{\partial K} = \frac{-(Q-c)-2\beta s\rho}{2} \cdot \frac{-2s(Q-c)}{(2K-\beta s^2)^2} > 0,$$

$$\frac{\partial q_{2R}}{\partial K} = \frac{\partial q_{2R}}{\partial \rho} \cdot \frac{\partial \rho}{\partial K} = \frac{Q-c+2\beta s\rho}{2} \cdot \frac{-2s(Q-c)}{(2K-\beta s^2)^2} < 0.$$

So, q_{2N} is the decreasing function of K , and q_{2R} is the increasing function of K .

When the scaling parameter is relatively large, the return rate is small. The small return rate implies that the quantity of remanufactured products in the second period is small. The firm therefore has to increase the price in the first period to maximize profits, which consequently leads to production decreasing. Moreover, since the total quantity is not affected by the investment, the amount of new products increases with the remanufactured products decreasing.

4 The multi-period acquisition and pricing model

In fact, due to the huge initial investment, the firm will invest in remanufacturing systems in stages. Hence, this section assumes that the firm has a planning horizon of M periods ($M > 2$). The model for the multi-period planning horizon is therefore given as:
(Model P2)

$$\max_{p_i, q_{iR}, \rho} \Pi = (p_1 - c)q_1 + \sum_{i=2}^M \beta^{i-1} [(p_i - c)q_{iN} + (p_i - c + s)q_{iR} - K_i \rho_i^2 / 2] \quad (7)$$

s.t.:

$$q_{iR} \leq \rho_i(Q - p_{i-1}), \quad i = 2, \dots, M, \quad (8)$$

$$K_{i-1} > K_i, \quad i = 3, \dots, M. \quad (9)$$

The definitions of the terms in P2 are similar to P1. Similarly, based on the KKT conditions, the optimal solution of P2 can be obtained in table 2.

TABLE 2 The optimal solution of P2

Policies	Period 1	Period i ($i=2, \dots, M-1$)	Period M
Pricing	$p_1 = \frac{Q+c-\beta s \rho_2}{2}$	$p_i = \frac{Q+c-\beta s \rho_{i+1}}{2}$	$p_M = \frac{Q+c}{2}$
Quantity	$q_1 = \frac{Q-c+\beta s \rho_2}{2}$	$q_{iN} = \frac{(Q-c)(1-\rho_i)+\beta s(\rho_{i+1}-\rho_i^2)}{2}$ $q_{iR} = \frac{\rho_i(Q-c+\beta s \rho_i)}{2}$	$q_{MN} = \frac{(Q-c)(1-\rho_M)-\beta s \rho_M^2}{2}$ $q_{MR} = \frac{\rho_M(Q-c+\beta s \rho_M)}{2}$
Acquisition			$\rho_i^* = \frac{s(Q-c)}{2K_i - \beta s^2}$

According to the expression of ρ_i^* , Proposition 1 is also valid in the multi-period case. With the improvement of the remanufacturing system, $K_{i-1} > K_i$. According to the principal of profit maximization, the cost savings are attractive to the firm. The firm has incentive to acquire and remanufacture more cores.

Hence, the investment increases with the passing of time ($I_{i-1} < I_i$). The optimal return rate also increases.

Proposition 3. (a) p_i ($i = 1, 2, \dots, M - 1$) is increasing with K_{i+1} , and the investment has no impact on p_M . (b) q_{iR} ($i = 2, \dots, M$) is decreasing with K_i . (c) q_{iN} ($i = 1, 2, \dots, M - 1$) is increasing with K_i , but decreasing with K_{i+1} . In addition, q_{MN} is increasing with K_M .

Proof: In period i ($i = 1, 2, \dots, M - 1$), the first order condition of p_i is:

$$\frac{\partial p_i}{\partial K_{i+1}} = \frac{\partial p_i}{\partial \rho_{i+1}} \cdot \frac{\partial \rho_i}{\partial K_{i+1}} = -\frac{s\beta}{2} \cdot \frac{-2s(Q-c)}{(2K_{i+1} - \beta s^2)^2} > 0, \text{ so } p_i \text{ is the increasing function of } K_{i+1}.$$

Similarly, the other conclusions can be obtained:
In period i ($i = 2, \dots, M$),

$$\frac{\partial q_{iR}}{\partial K_i} = \frac{\partial q_{iR}}{\partial \rho_i} \cdot \frac{\partial \rho_i}{\partial K_i} = \frac{Q-c+2\beta s \rho_i}{2} \cdot \frac{-2s(Q-c)}{(2K_i - \beta s^2)^2} < 0.$$

In period i ($i = 1, 2, \dots, M - 1$),

$$\frac{\partial q_{iN}}{\partial K_i} = \frac{\partial q_{iN}}{\partial \rho_i} \cdot \frac{\partial \rho_i}{\partial K_i} = -\frac{Q-c+2\beta s \rho_i}{2} \cdot \frac{-2s(Q-c)}{(2K_i - \beta s^2)^2} > 0,$$

$$\text{and } \frac{\partial q_{iN}}{\partial K_{i+1}} = \frac{\partial q_{iN}}{\partial \rho_{i+1}} \cdot \frac{\partial \rho_{i+1}}{\partial K_{i+1}} = \frac{\beta s}{2} \cdot \frac{-2s(Q-c)}{(2K_{i+1} - \beta s^2)^2} < 0.$$

Additionally, in period M ,

$$\frac{\partial q_{MN}}{\partial K_M} = \frac{\partial q_{MN}}{\partial \rho_M} \cdot \frac{\partial \rho_M}{\partial K_M} = -\frac{Q-c+2\beta s \rho_M}{2} \cdot \frac{-2s(Q-c)}{(2K_M - \beta s^2)^2} > 0$$

Due to $K_{i-1} > K_i$, it is easily concluded that the price decreases as more time periods pass. Therefore, the demand for the product increases. Since consumers cannot distinguish between new products and remanufactured products, and remanufactured products have larger marginal profit, the increasing remanufactured products will cannibalize the new products market. As time going on, the quantity of new products will decrease while the quantity of remanufactured products will increase.

5 Numerical examples

In this section, numerical examples are presented to validate the analysis conclusions and obtain some managerial insights. Firstly, in the two-period case, the parameters are set as follows: $Q = 10, c = 3.5, \beta = 0.95$,

$s = 1$, and $K \in [5, 23]$. The optimal policies under the different scaling parameters are showed in figure1-4. It is

obvious that Propositions 1 and 2 are justified by Figure1-4.

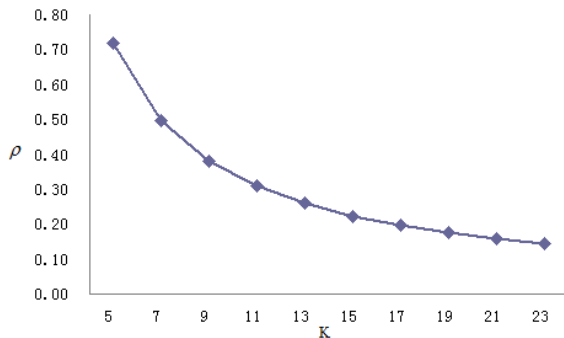


FIGURE 1 The optimal return rate

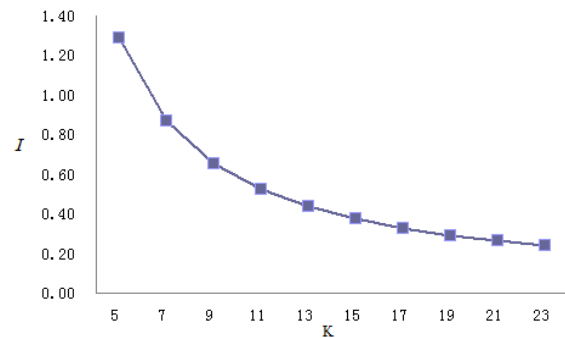


FIGURE 2 The initial investment

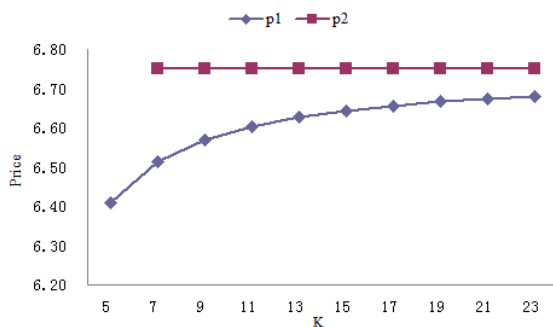


FIGURE 3 The optimal price

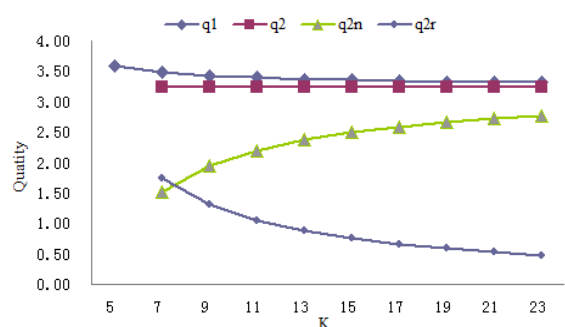


FIGURE 4 The optimal quantities

Next, the eleven-period case is considered. The parameters are set as follows: $Q = 10$, $c = 3.5$, $\beta = 0.95$, $s = 1$, and $K_i \in [4, 40]$. The result is shown in table 3. Table 3 shows that as the scaling parameter decreases from period 1 to 11, the respective optimal return rate

increases. Moreover, the quantity of remanufactured products and the investment in the remanufacturing system increases. With the exception of the last period, the price and quantity of new products decreases. These results cohere with the analytical conclusions.

TABLE 3 The optimal solution in the eleven-period case

Period	K_i	ρ_i^*	p_i	q_n	q_r	I_i	Π_i
1	-	-	7.4716	5.8608	-	-	14.4852
2	40	6.32%	7.4684	2.3718	0.1598	0.1598	20.7342
3	36	7.03%	7.4643	2.3576	0.1780	0.1780	26.9829
4	32	7.92%	7.4592	2.3399	0.2009	0.2009	33.2312
5	28	9.07%	7.4522	2.3172	0.2306	0.2306	39.4790
6	24	10.62%	7.4425	2.2871	0.2705	0.2705	45.7257
7	20	12.79%	7.4277	2.2453	0.3271	0.3271	51.9704
8	16	16.08%	7.4026	2.1838	0.4136	0.4136	58.2109
9	12	21.65%	7.3510	2.0868	0.5622	0.5622	64.4387
10	8	33.11%	7.1831	1.9397	0.8772	0.8772	70.5883
11	4	70.42%	7.5000	0.5163	1.9837	1.9837	76.8383

6 Conclusions

It is essential to consider the investment in remanufacturing systems when assessing the real situation of remanufacturing systems in developing countries. This paper formulated the two-period and multi-period model. Moreover, the relationship between the optimal policies and the scaling parameter was analysed. In addition, the numerical examples illustrated the analytical conclusions. In summary, this paper makes a contribution to the literature on remanufacturing by drawing attention to the real situation in developing countries. For future work, it is natural to consider the

optimal policies in a competitive market. The impact of government subsidy on remanufacturing also deserves attention, e.g., considering the effect of the constraint of lowest return rate regulated by the government.

Acknowledgments

This work was supported by National Social Science Foundation of China (No. 12CGL045), Fujian Provincial Natural Science Foundation of China (No. 2011J01381), China Postdoctoral Science Foundation (No. 2013M531541), and GuangXi Education Department Foundation of China (No. 2013YB146).

References

- [1] Hua G, Cheng T C E, Wang 2011 *International Journal of Production Economics* **132**(2) 178-85
- [2] Wang X, Li D, O'brien C 2010 *International Journal of Production Economics* **124**(2) 463-74
- [3] Kenne J P, Dejax P, Gharbi A 2012 *International Journal of Production Economics* **135**(1) 81-93
- [4] Georgiadis P, Athanasiou E 2010 *European Journal of Operational Research* **202**(2) 420-33
- [5] Vlachos D, Georgiadis P, Iakovou E 2007 *Computers & Operations Research* **34**(2) 367-94
- [6] Guide V D R, Teunter R H, Van Wassenhove L N 2003 *Manufacturing & Service Operations Management* **5**(4) 303-316
- [7] Pokharel S, Liang Y 2012 *International Journal of Production Economics* **38**(1) 170-76
- [8] Mutha A, Pokharel S 2009 *Computers & Industrial Engineering* **56**(1) 334-46
- [9] Lieckens K, Vandaele N 2012 *International Journal of Production Research* **50**(1) 23-40
- [10] Debo L G, Toktay L B, Van Wassenhove L N 2005 *Management Science* **51**(8) 1193-205
- [11] Heese H S, Cattani K, Ferrer G 2005 *European Journal of Operational Research* **164**(1) 143-57
- [12] Ferguson M E, Toktay L B 2006 *Production and operations management* **15**(3) 351-68
- [13] Ferrer G, Swaminathan J M 2006 *Management Science* **52**(1) 15-26
- [14] Ferrer G, Swaminathan J M 2010 *European Journal of Operational Research* **206**(2) 370-9
- [15] Scotto Webster, Supriya Mitra 2007 *Journal of Operation Management* **25**(6) 1123-40
- [16] Supriya Mitra, Scott Webster 2008 *International Journal of Production Economics* **119**(2) 287-98
- [17] Savaskan R C, Bhattacharya S, Van Wassenhove L N 2004 *remufacturing Management Science* **50**(2) 239-52

Authors	
	<p>Yang Chenghu, born in January, 1980, Fuzhou County, Fujian Province, PR China</p> <p>Current position, grades: Associate Professor of All-Trans Logistics School, Fuzhou University, PR China. University studies: D.Sc.Tech. from Southwest Jiaotong University in PR China. Scientific interest: supply chain management and inventory management. Publications: more than 20 papers published in various journals. Experience: teaching experience of 6 years, 6 scientific research projects.</p>
	<p>Liu Lanying, born in May, 1990, Fuzhou County, Fujian Province, PR China</p> <p>Current position, grades: Master Student of All-Trans Logistics School, Fuzhou University, PR China. University studies: B.Sc. in Logistics Management from Fuzhou University in PR China. Scientific interest: supply chain management and inventory management.</p>
	<p>Zhang Lei, born in January, 1978, Nanning County, Guangxi Province, PR China</p> <p>Current position, grades: Associate Professor of College of Economics and Management, Guangxi Teachers Education University, PR China. University studies: D.Sc.Tech. from Southwest Jiaotong University in PR China. Scientific interest: supply chain management. Publications: more than 20 papers published in various journals. Experience: teaching experience of 4 years, 2 scientific research projects.</p>

Continuous-time optimal portfolio model with mean-reverting process

Xing Yu*

Department of Mathematics & Applied Mathematics Hunan University of humanities, science and technology, Loudi, 417000, P.R. China

Received 1 March 2014, www.tsi.lv

Abstract

This paper studies a continuous-time portfolio optimization problem. It is proposed a simple but powerful approximation approach that is both accurate and computationally efficient for the terminal expectation of the investors with mean-reverting process, which is different from the existing literatures that apply the dynamic programming method. Numerical examples illustrate the computational efficiency and accuracy of our approach when compared with results from Monte Carlo (MC) simulations.

Keywords: Continuous-time portfolio, Mean-reverting process, Optimization; Monte Carlo

1 Introduction

It was pioneered by Markowitz to propose the mean-variance(M-V) framework, which has been playing a cornerstone role in the theory of portfolio selection [1]. Numerous scholars have extended Markowitz's single period portfolio selection into the counterparts of multi-period and continuous-time. Mossin (1968) [2] considered optimal multi-period portfolio selection by using dynamic programming approach. Hakansson (1971) [3] presented the multi-period mean-variance model using a general portfolio theory. Wu (2014) [4] considered a multi-period mean-variance portfolio selection when the time horizon is assumed to be stochastic and depends on the market states. For more detailed discussion on the subject of dynamic portfolio selection, it is referred to [5-8]. In continuous-time version of dynamic portfolio selection, Yao [9] investigated a continuous-time mean-variance portfolio selection problem with multiple risky assets using the Lagrange duality method and the dynamic programming approach, and derived explicit closed-form expressions of efficient frontier. Yong [10] considered a continuous-time optimal consumption and portfolio selection model with voluntary retirement using the dynamic programming method to derive the optimal strategies in closed-form. Holger [11] studied constrained portfolio problems and solved the problems by dynamic programming. The standard approach to solve the dynamic portfolio optimization problem is martingale method, which was developed by Karatzas et.al (1987) [12] and Cox and Huang (1989) [13]. Because dynamic programming method to solve the problem gives easy access to the value function, many scholars pay much attention to it.

To the best of our knowledge, all the existing literatures about continuous-time portfolio selection

model are solved by analysed method, there has been few literatures on approximation method. In fact, for certain investment, all the strategies are not accurate, because the finance market-self is not certain. Therefore, it is reasonable to use approximation method to solve the problem. To deal with uncertain, numerous research on fuzzy mathematic method [14-15]. This paper proposes a mean-reverting process to approximate the distribution of a weighted sum of correlated assets.

The paper is organized as follows. Section 2 gives model formulations, including the mean-reverting process and the two steps of approximation. Section 3 describes the optimal portfolio selection model. Numerical examples, illustrating both the computational efficiency and accuracy of our method are presented in section 4. Section 5 gives concluding remarks.

2 Model formulations

It is considered a portfolio consisting of m assets with price $S_i(t), i=1,2,\dots,m$, which are described as the following stochastic differential equation,

$$\begin{cases} dS_i(t) = \mu_i(t)S_i(t)dt + \sigma_i(t)S_i(t)dz_i t, \forall t \in [0, T], \\ S_i(0) = s_{i0} \end{cases} \quad (1)$$

where $\{z_i(t), t > 0\}$ are standard Brownian motions, and $\mu_i(t)$, $\sigma_i(t)$ are respectively the appreciation rate and vitality rate of asset i . It is assumed that $\text{cov}(dz_i, dz_j) = \rho_{ij}dt$, ρ_{ij} denote the correlation coefficients between $z_i(t)$ and $z_j(t)$.

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

The portfolio via $S(\cdot)$ can be expressed by $\sum_{i=1}^n \omega_i S_i(t)$, ω_i is the weight of asset i . As we know, a weighted sum of non-independent lognormals does not have an explicit analytical expression. One can use Monte Carlo simulation techniques to obtain a numerical estimation. Other techniques take account in approximating. For instance, it is used the lognormal distribution for approximating the sum of lognormals. But it leads poor results. Another different methods can see to Albrecher (2004), Curran (1994), and Kawai (2003) [16-18].

In this paper, we present an approximate method of mean-reverting prices. Therefore, it is necessary to introduce it firstly.

2.1 MEAN-REVERTING PROCESS

According to Bos [19], the process x_i follows the exponential Ornstein – Uhlenbeck (EOU) model

$$\begin{cases} dx_i(t) = k_i(t)(\gamma_i(t) - x_i(t))dt + v_i(t)x_i(t)dz_i(t), \forall t \in [0, T], \\ x_i(0) = x_{i0} \end{cases} \quad (2)$$

where the parameters $v_i(t)$, $\gamma_i(t)$ and $k_i(t)$ are determined following Duffie and Richardson [20].

A direct calculation yields the first and second moments of $x_i(t)$ are

$$Ex_i(t) = e^{-\int_0^t 2k_i(s)ds} \left(x_{i0} + \int_0^t k_i(s)\gamma_i(s) e^{\int_0^s k_i(u)du} ds \right), \quad (3)$$

$$Ex_i^2(t) = e^{\int_0^t (v_i^2(s) - 2k_i(s))ds} \int_0^t v_i^2(s) \left(x_{i0} + \int_0^s k_i(u)\gamma_i(u) e^{\int_0^u k_i(v)dv} du \right)^2 e^{-\int_0^s v_i^2(v)dv} ds + e^{-\int_0^t 2k_i(s)ds} \left(x_{i0} + \int_0^t k_i(s)\gamma_i(s) e^{\int_0^s k_i(u)du} ds \right)^2 \quad (4)$$

Let the portfolio $X(t) = \omega X(t)$, and suppose it follows a mean-reverting process, and let $x_{i0} = S_{i0}$. It means that the mean at the time points of small intervals is with reverting.

2.2 THE APPROXIMATION IN THE FIRST STEP

Suppose the assets prices follow (1), in every small interval $[t, t + \Delta t]$, $l = 1, 2 \dots n$.

$$\begin{cases} \mu_i(t) = \mu_{il}, \forall t \in [t_{l-1}, t_l], l = 1, 2 \dots n. \\ \sigma_i(t) = \sigma_{il} \end{cases} \quad (5)$$

To calculate of (1) in $[t_{l-1}, t_l]$ yields

$$\begin{cases} ES_i(t_l) = ES_i(t_{l-1}) e^{\mu_{il} \Delta t_l} \\ ES_i^2(t_l) = ES_i^2(t_{l-1}) e^{(\sigma_{il}^2 + 2\mu_{il}) \Delta t_l} \end{cases} \quad (6)$$

So

$$\begin{cases} \mu_{il} = \frac{1}{\Delta t_l} \ln \frac{ES_i(t_l)}{ES_i(t_{l-1})} \\ \sigma_{il} = \sqrt{\frac{1}{\Delta t_l} \left[\ln \frac{ES_i(t_l)^2}{ES_i(t_{l-1})^2} - 2 \ln \frac{ES_i(t_l)}{ES_i(t_{l-1})} \right]} \end{cases} \quad (7)$$

The first step approximation is that $S_i(\cdot) \approx X(\cdot)$ in every sufficiently small interval $[t_{l-1}, t_l]$, and let the first and second moments of $S_i(t)$ is the same as $x_i(t)$ at point t_{l-1} and t_l . We obtain

$$\begin{cases} \mu_{il} = \frac{1}{\Delta t_l} \ln \frac{Ex_i(t_l)}{Ex_i(t_{l-1})} \\ \sigma_{il} = \sqrt{\frac{1}{\Delta t_l} \left[\ln \frac{Ex_i(t_l)^2}{Ex_i(t_{l-1})^2} - 2 \ln \frac{Ex_i(t_l)}{Ex_i(t_{l-1})} \right]} \end{cases} \quad (8)$$

2.3 THE APPROXIMATION IN THE SECOND STEP

The second step approximation is $X(t) \approx \sum_{i=1}^m \omega_i S_i(t)$. A direct calculation yields

$$\begin{cases} EX(T) \approx \sum_{i=1}^m \omega_i x_{i0} \exp\left(\sum_{l=1}^m \mu_{il} \Delta t_l\right) \\ EX^2(T) \approx \sum_{i=1}^m \sum_{j=1}^m \omega_i \omega_j x_{i0} x_{j0} \exp\left(\sum_{l=1}^m (\mu_{il} + \mu_{jl} + \rho_{ij} \sigma_{ij}) \Delta t_l\right) \end{cases} \quad (9)$$

To classify the proposed approximation method, it is summarized the procedure as follows:

- Divide the time period $[0, T]$ into finitely many small intervals $[t_{l-1}, t_l]$ and calculate the first and second moments according to (3), (4).
- Substitute the first and second moments to (8), yield μ_{il} and σ_{il} .
- Substitute μ_{il} and σ_{il} to (2.9) obtain the first and second moments of the approximated portfolio at the end of the investment horizon.

3 The optimal portfolio selection model

In order to show the effectiveness, we only consider the simply case. The problem for an investor is to find the

optimal strategy to minimize the variance while attaining a given level of the expected wealth.

$$\begin{cases} \min_{\omega} \text{Var}(X(T)) \\ \text{s.t. } E(X(T)) = u \end{cases} \quad (M1)$$

where u is a pre-given constant, representing the expected value level, which the investor requires to achieve and $\text{Var}(X(T)) = E(X - EX)^2 = EX^2(T) - (EX(T))^2$, $EX^2(T)$, $EX(T)$ are obtained from (9).

According to the optimal portfolio model (M1), the problem can be deal with by the Lagrange method. Introducing the Lagrange multiplier λ leads to the following problem

$$\min E(X(T) - u)^2 + 2\lambda(EX(T) - u). \quad (M2)$$

Let $\pi(T) = g(x(T), \lambda)$ be the optimal solution of the Lagrangian problem (M2) and $G(x_0, \lambda)$ be the optimal value. According to the Lagrange duality theory, if λ^* satisfies $\max_{\lambda} G(x_0, \lambda)$, then $\pi^*(t) = g(x(t), \lambda^*)$ is the optimal shares of (M1) and $G(x_0, \lambda^*)$ is its optimal value. Problem (M2) is equivalent to $\min E(x(T) + a)^2$, where $a = \lambda - u$.

4 Numerical examples and applications

In this subpart, we illustrate the accuracy and computational efficiency from two-asset and four assets experiments, and compare our results with MC simulations.

- Two-asset Case

Time to maturity $T = 1$ year, risk-free rate 5% per annum, $v_1 = v_2 = 20\%$, $S_{10} = x_{10} = 50$, $S_{20} = x_{20} = 50$, $\rho = 0.5$, $\gamma_1 = x_{10}e^{5\%T} = 52.5636$, also so, $\gamma_2 = x_{20}e^{5\%T} = 52.5636$, $k_1 = k_2 = \frac{1}{3}$, $u = 8\%$.

Reference

[1] Markowitz H 1952 Portfolio selection *Journal of Finance* 7 77-91
 [2] Mossion J 1968 Optimal multi-period portfolio policies *Journal of Business* 5 215-29
 [3] Huiling Wu, Yan Zeng, Haxiang Yao 2014 Multi-period Markowitz's mean-variance portfolio selection with state-dependent exit probability *Economic Modelling* 36 69-78
 [4] Costa O L V, Araujo M V 2008 A generalized multi-period mean-variance portfolio optimization with Markov switching parameters *Automatica* 44(10) 2487-97
 [5] Basak S, Chabakauri G 2010 Dynamic mean-variance asset allocation *Review of Finance Study* 23 2970-3016

Following the step of part 2, solving (M1), the Monte Carlo simulation results of portfolio weights is $\omega_1 = 0.3891, \omega_2 = 0.6109$, and the results from the approximation method proposed in our study is $\omega_1 = 0.3255, \omega_2 = 0.6755$.

- Four-asset Case

Time to maturity $T = 1$ year, risk-free rate 5% per annum, $v_1 = v_2 = v_3 = v_4 = 20\%$, $S_{10} = x_{10} = 50$, $S_{20} = x_{20} = 50$, $S_{30} = x_{30} = 50$, $S_{40} = x_{40} = 50$, $k_1 = k_2 = k_3 = k_4 = \frac{1}{3}$, $u = 8\%$.

Following the step of part 2, solving (M1), the Monte Carlo simulation results of portfolio weights is $\omega = 0.0414, \omega = 0.2782, \omega = 0.6351, \omega = 0.0453$, and the results from the approximation method proposed in our study is $\omega_1 = 0.0527, \omega_2 = 0.3181, \omega_3 = 0.5324, \omega_4 = 0.0968$.

From the compared results, we find that the portfolio shares from our model and Monte Carlo simulation is nearly the same.

5 Conclusion

Different from the general dynamic approach, we propose the mean-reverting process to approximate the weight sum of assets, and extend the discrete portfolio to the continuous-time case. The optimal portfolio selection model aims to minimize the risk described as VaR meanwhile to get the given expectation is presented. Numerical example results for two-asset and four-asset case shows the accuracy and computational efficiency compared to Monte Carlo simulation results. It provides a new way to deal with the continuous-time portfolio selection.

Acknowledgment

This research is supported by the key project of hunan province department of education (12A077).

[6] Glaounouridis D, Vrontos I D 2007 Hedge fund portfolio construction: a comparison of static and dynamic approaches *Journal of banking and finance* 31 199-217
 [7] Wu H 2014 Multi-period Markowitz's mean-variance portfolio selection with state-dependent exit probability *Economic Modelling* 36 69-78
 [8] Yao H X 2014 Continuous-time mean-variance portfolio selection with only risky assets *Economic modelling* 36 244-51
 [9] Yong H S 2012 Voluntary retirement and portfolio selection: Dynamic programming approaches *Applied mathematics letter* 25 1087-93
 [10] Holger K, Mogens S A 2013 dynamic programming approach to constrained portfolios *European Journal of operational Research* 229 453-61

- [11] Karatzas I, Lehoczky J P, Shreve S E 1987 Optimal portfolio and consumption decisions for a small investor on a finite horizon *SIAM Journal of control and optimization* **25** 1557-86
- [12] Cox J C, Huang C F 1989 Optimal consumption and portfolio policies when asset prices follow a diffusion process *Journal of economic theory* **49** 33-53
- [13] Ostermark R 1996 A fuzzy control model (FCM) for dynamic portfolio management *Fuzzy sets and systems* **78**(3) 243-54
- [14] Yong J L 2012 Fuzzy multi-period portfolio selection optimization models using multiple criteria *Automatica* **48** 3042-53
- [15] Albrecher H, Predota M 2004 On Asian option pricing for NIG levy processes *Journal of computational and applied mathematics* **172** 153-68
- [16] Valuing C M 1994 Asian and portfolio options by conditioning on the geometric mean price *Management science* **40** 1705-11
- [17] Kawai A 2003 A new approximate swaption formula in the LIBOR market model: an asymptotic expansion approach *Applies mathematical finance* **10** 49-74
- [18] Bos L P A, Ware A F, Pavlov B S 2002 On a semi-spectral method for pricing an option on a mean-reverting asset *Quantitative Finance* **2** 337-45
- [19] Duffie D, Richardson H R 1991 Mean-variance hedging in continuous time *The Annals of Applied Probability* **1** 1-15

Authors



Xing Yu, born on February 15, 1981, in Xianning City of Hubei province

Current positions, grades: Hunan university of humanities, Science and technology (China); Lector

Professional interests: Applicate mathematics; Finance model

Research interests: the optimal portfolio model; mathematical model

Evolutionary game analysis of enterprises' technological innovation strategies

Nannan Duan^{1, 2*}, Fuyuan Xu¹, Ming Ni²

¹Business School, University of Shanghai for Science & Technology, Shanghai, China, 200093

²School of Economics and Management, East China JiaoTong University, Nanchang, China, 330013

Received 19 January 2014, www.tsi.lv

Abstract

Different from general studies on competition and/or cooperation relationship of enterprises, the paper classified enterprises' technological innovation strategies into cooperation, neutrality and competition, and analysed the evolution of enterprises' relationship and strategy selection in technological innovation using the evolutionary game method and the MATLAB simulation technique. The paper drew the following conclusions: (1) the increase in technological content may cause two uncertain cases of enterprises' strategy, namely both parties chose the cooperation strategy, or one party chose the cooperation strategy while the other party chose the noncooperation strategy; (2) the increase in innovation revenue (coefficient) prompted enterprises to eventually tend towards the cooperative innovation strategy, no matter what the initial relationship between enterprises was; (3) the increase in the number of enterprises with the neutrality attitude promoted enterprise groups to tend towards cooperative innovation.

Keywords: technological innovation, evolutionary game, cooperation, neutrality, competition

1 Introduction

As a key for global competitiveness and business success, innovation is a critical element for boosting economies and societal development. Only those firms that foster innovation and handle innovation processes properly can survive eventually in the global market [1, 2]. Many scholars and managers have studied the technological innovation from various perspectives. Yuyin Yi et al. studied the opportunistic behaviors affecting the smooth development of cooperative R&D of enterprises by building an evolutionary game model, and proposed to solve such behaviours by means of supervision [3]. PRASHANT et al. believed that the higher the degree of openness between both parties in cooperation was, the more knowledge both parties acquired from the alliance; meanwhile, as the openness increased, it was more possible for cooperative partner to adopt opportunistic behaviours and cause more harms [4]. In order to find the operational mechanism of inter-enterprise cooperation, Zhaode Liu and Weiguo Zhang explored how enterprises chose the cooperative enterprise and cooperative innovation way by analysing the innovative behaviours of new high-tech enterprises [5]. Cristina and Carlos identified alternative strategic behaviours from the combination of competitive and cooperative attitudes and analysed the effect of co-opetitive strategy on technological innovation [6]. Liang Xu et al. believed that the competition & cooperation strategy of enterprises could promote technological innovation performance

significantly and the cooperation strategy could promote technological innovation of enterprises greatly [7].

Unfortunately, as we shall highlight, many previous empirical studies exploring the technology innovation based on competition and/or cooperation relationship do not consider another relationship that is neither competition nor cooperation, namely neutrality. The paper introduces the view into the analysis on the evolutionary process of technological innovation, and believes that inter-enterprise relationship in technological innovation process includes the cooperation and the noncooperation, and the noncooperation includes the competition and the neutrality. In addition, the innovation is a process of technological complementation and benefit redistribution, so choosing which strategy depends on how the enterprise weighs the revenue from technological innovation, the risk loss caused by the potential opportunistic behaviour of cooperative partner, and the cost of technological innovation. Therefore, the paper chooses technological content, technological innovation revenue coefficient and ratio of neutrality as the impact factors in the analysis, and establishes a model of enterprises' technological innovation using the evolutionary game theory to make an in-depth analysis on the important factors affecting inter-enterprise relationship and various valuable cases of game evolution.

* *Corresponding author* e-mail: crossrainbow81@163.com

2 Evolutionary model of technological innovation

The evolutionary game theory is based on biological evolutionism’s basic principle - the survival of the fittest, and combines the game theory analysis with the dynamic evolution process analysis to reflect the process of dynamic strategy adjustment and equilibrium selection of participants [8]. Because there are information incompleteness and asymmetry in the technological innovation process, when deciding which strategy to adopt, the enterprise may use other enterprise groups’ strategies for reference and adjust its strategy by copying other enterprises and using the trail-and-error method, and eventually tends to be stable. Therefore, the evolutionary game method is applicable to the study on the characteristics of inter-enterprise technological innovation strategy selection and corresponding evolutionary development.

2.1 HYPOTHESES OF MODEL

Hypothesis 1: Abstract some enterprises producing homogeneous products into two enterprise groups: A and B. As mentioned above, in the technological innovation, the two enterprise groups have two strategic choices: cooperation and noncooperation. Assume A has a probability of x to adopt noncooperation, and then it has a probability of $1-x$ to adopt cooperation; similarly, B has a probability of y to adopt noncooperation, and then it has a probability of $1-y$ to adopt cooperation.

Hypothesis 2: In the case that both groups adopt noncooperation, enterprises have different motives. That is, some enterprises may adopt noncooperation in the sense of competition due to an opportunistic motive or the overprotection to technologies; while, other enterprises may hold neutrality because of their large technological distance with other enterprises or for the purpose of high risk aversion. Assume the rates of neutrality in A and B are both expressed as z , and the rates of competition are both expressed as $1-z$.

Hypothesis 3: Table 1 shows the assumptions of the expression, meaning and value range of some coefficients necessary for model analysis.

TABLE 1 Relative Symbols and Their Meanings and Value Ranges

Symbol	Meaning	Value Range
q_1	Success rate of independent innovation	$0 \leq q_1 \leq 1$
q_2	Success rate of innovation with opportunistic behaviour	$0 \leq q_2 \leq 1$
q_3	Success rate of cooperative innovation	$0 \leq q_3 \leq 1$
α	Coefficient of technological innovation revenue	$\alpha > 1$
W	Technological content (difficulty) of technological innovation project	$W \geq 0$
λ_1	Cost coefficient of competition	$\lambda_1 \geq 1$
λ_2	Cost coefficient of opportunism	$\lambda_2 \geq 1$
λ_3	Cost coefficient of cooperative innovation	$\lambda_3 \geq 1$
λ_4	Cost coefficient of neutral innovation	$\lambda_4 \geq 1$
k	Competition index	$k > 1$

Hypothesis 4: For simplicity, assume enterprise revenue has a linear relation of $q_i\alpha W$ ($i=1,2,3$) with the successful rate of technological innovation and technological content. Then, under the condition of technological content W , if both parties adopt noncooperation, they get the revenue of $q_1\alpha W$; if one party adopts neutrality while the other party adopts cooperation, two parties still accomplish their innovations independently because they fail to achieve cooperative innovation eventually, and exactly for this reason, they still both get the revenue of $q_1\alpha W$; if one party seems to cooperate but actually adopts competition while the other party adopts cooperation, there will be an opportunistic phenomenon - the party of competition may act as a free rider. In this case, both parties may get the revenue of $q_2\alpha W$ from this form of cooperation; when both parties adopt cooperative innovation, their revenues are both assumed as $q_3\alpha W$.

Hypothesis 5: Assume the total cost of technological innovation changes as the enterprise’s innovation strategy changes.

(1) When both parties adopt competition, or one party keeps neutrality while the other party adopts competition, with the certain technological content W , the competition enterprise’s initial cost input is basically in direct proportion to its technological content (approaching a linear function). As the technological innovation goes deeper, the enterprise gives more positive comments on the technology’s importance and future prospect, thus increasing cost input. When the conditions are fitted, we find the relations of elements more approach an exponential function. Therefore, assume the cost as $c_1 = \lambda_1 W^k$.

(2) When one party adopts cooperative innovation while the other party adopts competition, the party of competition shows an opportunistic behaviour, trying to learn knowledge from the other party and terminating the cooperation after acquiring the technology. So under the condition of certain technological content W , the enterprise’s initial cost input is basically in direct proportion to its technological content (approaching a linear relation). However, as the technological innovation goes deeper, cost input doesn’t change as technological content changes, and is about to reach a stable value. Therefore, assume the cost as $c_2 = \lambda_2 \ln W$.

(3) When both parties adopt the cooperative innovation, the cost input is in direct proportion to technological content, namely $c_3 = \lambda_3 W$.

(4) When both parties or one party adopts neutrality, the innovation cost of neutral enterprise is $c_4 = \lambda_4 W$.

2.2 EVOLUTIONARY GAME MODEL

According to the hypotheses above, Table 2 shows the payoff matrix (including 9 cases) of the technological innovation mode selections of A and B.

TABLE 2 Pay-off matrixes of Evolutionary Game of A and B (include 9 cases)

Enterprise Group A \ Enterprise Group B		Noncooperation with a Probability of y		
		Neutrality with a Probability of z	Competition with a Probability of $1-z$	Cooperative Innovation with a Probability of $1-y$
Noncooperation with a Probability of x	Neutrality with a Probability of z	$(q_1\alpha W - \lambda_4 W, q_1\alpha W - \lambda_4 W)$	$(q_1\alpha W - \lambda_4 W, q_1\alpha W - \lambda_1 W^k)$	$(q_1\alpha W - \lambda_4 W, q_1\alpha W - \lambda_3 W)$
	Competition with a Probability of $1-z$	$(q_1\alpha W - \lambda_1 W^k, q_1\alpha W - \lambda_4 W)$	$(q_1\alpha W - \lambda_1 W^k, q_1\alpha W - \lambda_1 W^k)$	$(q_2\alpha W - \lambda_2 \ln W, q_2\alpha W - \lambda_3 W)$
Cooperative Innovation with a Probability of $1-x$		$(q_1\alpha W - \lambda_3 W, q_1\alpha W - \lambda_4 W)$	$(q_1\alpha W - \lambda_3 W, q_1\alpha W - \lambda_4 W)$	$(q_2\alpha W - \lambda_3 W, q_2\alpha W - \lambda_2 \ln W)$

Table 2 can be simplified and sorted into Table 3 in which the payoff matrix includes 4 cases. Table 3 shows that when both parties adopt noncooperation, each of them get the revenue of D ; when one party adopts cooperative innovation while the other party adopts

noncooperation, the party of noncooperation gets the revenue of E , and the party of cooperative innovation gets the revenue of F ; when both parties adopt cooperative innovation, they both get the revenue of G .

TABLE 3 Payoff Matrix of Evolutionary Game of A and B (include 4 cases)

Enterprise Group A \ Enterprise Group B	Noncooperation with a Probability of y	Cooperative Innovation with a Probability of $1-y$
Noncooperation with a Probability of x	(D, D)	(E, F)
Cooperative Innovation with a Probability of $1-x$	(F, E)	(G, G)

Apparently, the expressions of D , E , F and G in Table 3 can be obtained by substituting corresponding contents in Table 2. The expressions are as follows:

$$D = q_1\alpha W - (1-z)\lambda_1 W^k - z\lambda_4 W, \tag{1}$$

$$E = z(q_1\alpha W - \lambda_4 W) + (1-z)(q_2\alpha W - \lambda_2 \ln W), \tag{2}$$

$$F = \alpha Wz(q_1 - q_2) + q_2\alpha W - \lambda_3 W, \tag{3}$$

$$G = q_3\alpha W - \lambda_3 W. \tag{4}$$

According to Table 3, U_{11} , U_{12} and \bar{U}_1 , which are the fitness functions and the average fitness function when A adopts noncooperation and cooperation, can be expressed as:

$$U_{11} = Dx + E(1-x), \tag{5}$$

$$U_{12} = Fx + G(1-x), \tag{6}$$

$$\bar{U}_1 = xU_{11} + (1-x)U_{12}. \tag{7}$$

Similarly, when B adopts noncooperation and cooperation, the fitness functions and the average fitness functions U_{21} , U_{22} , \bar{U}_2 are respectively:

$$U_{21} = Dx + E(1-x), \tag{8}$$

$$U_{22} = Fx + G(1-x), \tag{9}$$

$$\bar{U}_2 = yU_{21} + (1-y)U_{22}. \tag{10}$$

According to the duplicator dynamic equation, the following two-dimensional differentiable dynamic system can be obtained:

$$\begin{cases} \frac{dx}{dt} = x(U_{11} - \bar{U}_1) = x(1-x)[(D-E-F+G)y + E-G] \\ \frac{dy}{dt} = y(U_{21} - \bar{U}_2) = y(1-y)[(D-E-F+G)x + E-G] \end{cases} \tag{11}$$

Then, let $\frac{dx}{dt} = \frac{dy}{dt} = 0$, and get the following five equilibrium points of equation set: $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, (x_0, y_0) , among which $x_0 = y_0 = \frac{G-E}{D-E-F+G}$ ($x_0, y_0 \in [0,1]$).

After calculation, the Jacobian matrix J is:

$$J = \begin{bmatrix} (1-2x)[(D-E-F+G)y+E-G] & x(1-x)(D-E-F+G) \\ y(1-y)(D-E-F+G) & (1-2y)[(D-E-F+G)x+E-G] \end{bmatrix}. \tag{12}$$

According to the stability theory of nonlinear differential equation, the stability of balance points can be determined by the sign of the Jacobian matrix's characteristic root [9]. Table 4 shows the values of

matrix J 's determinant $DetJ$ and trace trJ in equilibrium points and other related information.

TABLE 4 Evolution Path of enterprise's technological innovation strategy selection

Condition	Balance Point	Sign of $DetJ$	Sign of trJ	Local Stability
	(0,0)	+	-	ESS
case1: $D-F > 0, G-E > 0$ Note: there are five balance points	(0,1)	+	+	Unstable point
	(1,0)	+	+	Unstable point
	(1,1)	+	-	ESS
	(x_0, y_0)	-	0	saddle point
Case 2: $D-F < 0, G-E > 0$ Note: there are four balance points, and $(x_0, y_0) \notin [0,1]$	(0,0)	+	-	ESS
	(0,1)	-	\pm	saddle point
	(1,0)	-	\pm	saddle point
	(1,1)	+	+	Unstable point
Case 3: $D-F > 0, G-E < 0$ Note: there are four balance points, and $(x_0, y_0) \notin [0,1]$	(0,0)	+	+	Unstable point
	(0,1)	-	\pm	saddle point
	(1,0)	-	\pm	saddle point
	(1,1)	+	-	ESS
Case 4: $D-F < 0, G-E < 0$ Note: there are five balance points	(0,0)	+	+	Unstable point
	(0,1)	+	-	ESS
	(1,0)	+	-	ESS
	(1,1)	+	+	Unstable point
	(x_0, y_0)	-	0	saddle point

According to the formulations (1) to (4), the following formulations can be obtained:

$$D-F = \alpha W(q_1 - q_2)(1-z) - (1-z)\lambda_1 W^k - z\lambda_4 W + \lambda_3 W, \tag{13}$$

$$G-E = q_3 \alpha W - \lambda_3 W - z(q_1 \alpha W - \lambda_4 W) - (1-z)(q_2 \alpha W - \lambda_2 \ln W). \tag{14}$$

3 Numerical simulation of evolutionary game model

3.1 ANALYSIS ON SYSTEM'S ASYMPTOTIC STABILITY AND PARAMETER SENSITIVITY

To make the simulation analysis, we need to assign values to related parameters. Assume A's probability to adopt noncooperation $x=0.3$, B's probability to adopt noncooperation $y=0.7$, the success rate of technological innovation in the case of noncooperation $q_1=0.4$, the success rate of technological innovation in the case of opportunistic behaviour $q_2=0.5$, the success rate of technological innovation in the case of cooperative innovation $q_3=0.8$, the competition index of in the case of competition strategy $k=1.5$, the cost coefficient of technological innovation in the case of competition

$\lambda_1=6$, the cost coefficient of technological innovation in the case of opportunistic behaviour $\lambda_2=4$, the cost coefficient of technological innovation in the case of cooperation $\lambda_3=5$, the cost coefficient of technological innovation in the case of neutrality $\lambda_4=4$, and the probability of neutrality in the case of noncooperation $z=0.4$.

According to the discussion above and values assignment to parameters, the next section further analyzes each case.

Case 1: To meet conditions of $D-F > 0$ and $G-E > 0$, assume the coefficient of technological innovation revenue $\alpha=8$ and the technological content of technological innovation $W=4$. In this case, the system's evolutionary phase diagram and evolution path are shown in fig.1a and fig.1b. In the figures, A and B

finally become stable in points (0,0) and (1,1) after a period of game. The system has two stable points, so which one is the final stable point depends on the initial values of x and y . That is, in the context that most enterprises in the market tend to adopt cooperative innovation, people are guided to adopt cooperative innovation eventually. However, when most enterprises in the market adopt noncooperation, both A and B eventually adopt noncooperation after a period of game, just as shown in fig.1b.

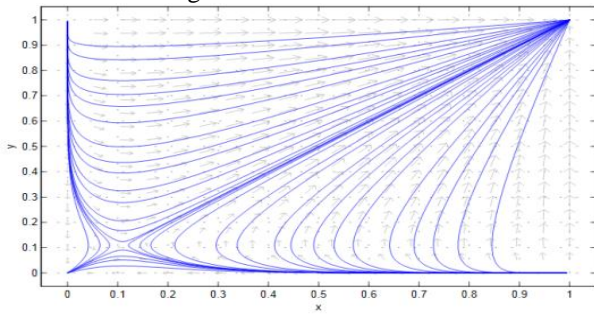


FIGURE 1a Phase Diagram in case 1

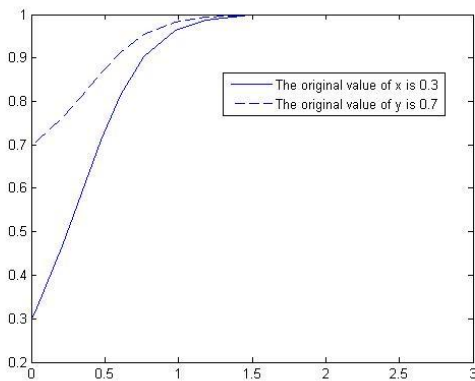


FIGURE 1b Evolutionary Path with stable point (1,1)

Adjust the value of W , α and z respectively, the game strategies of A and B may evolve correspondingly.

(1) Adjust the initial value of technological content W . Make $W = 4, 6, 10, 14, 16$ in turn. Fig.2a and fig.2b show that the stable point of the game between A and B changes from point (1,1) to point (0,0), and when the number of W 's value exceeds 16, the overall case evolves into case 2 with only one stable point (0,0).

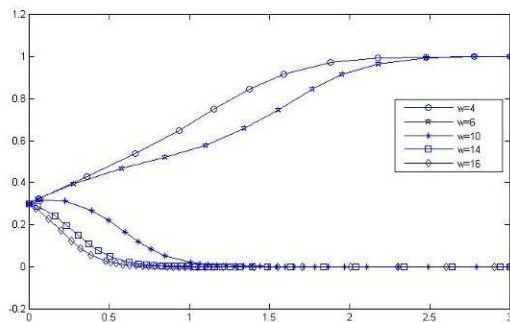


FIGURE 2a A's evolutionary Path figure as W increase

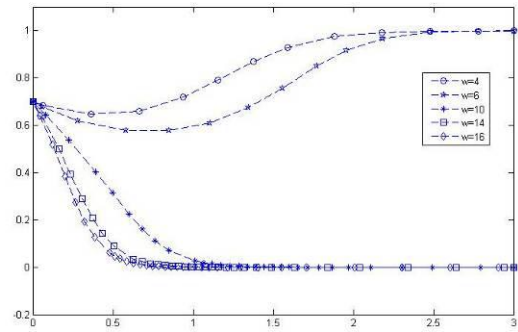


FIGURE 2b B's evolutionary Path figure as W increase

(2) Adjust the initial value of α . Make $\alpha = 8, 14, 20, 25, 28$ in turn. Fig.3a and fig.3b show that the stable strategy of A and B also changes from point (1,1) to point (0,0). It means that the enterprise is willing to turn from independent innovation to cooperation as technological innovation revenue increases.

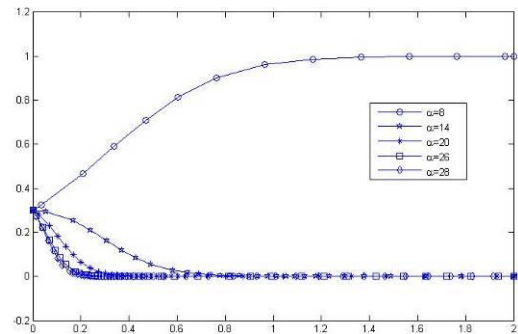


FIGURE 3a A's evolutionary Path figure as α increase

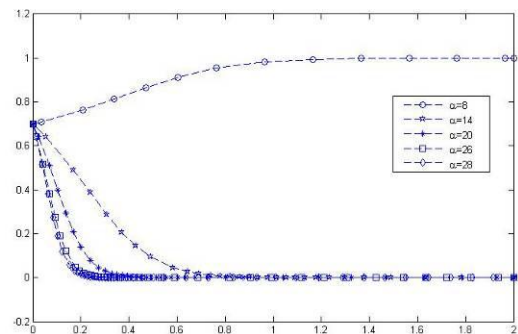


FIGURE 3b B's evolutionary Path figure as α increase

(3) Change the value of z . Make $z = 0.4, 0.5, 0.6, 0.7, 0.8$ in turn. Fig.4a and fig.4b show that as the number of enterprises of neutrality increases, the stable strategy of A and B shifts from (1,1) to (0,0) gradually. The phenomenon has important significance. It indicates that the increase in the number of neutrality may help to form inter-enterprise cooperation under certain conditions.

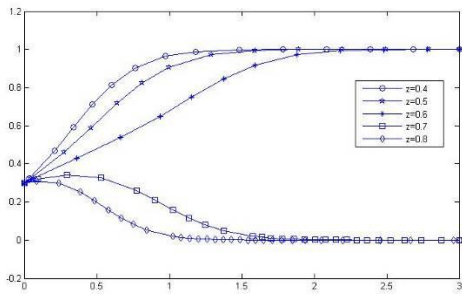


FIGURE 4a A's evolutionary Path figure as z increase

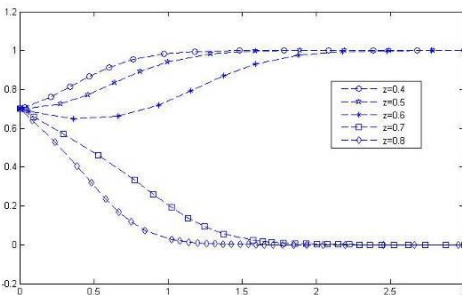


FIGURE 4b B's evolutionary Path figure as z increase

Case 2: To meet conditions of $D - F < 0$ and $G - E > 0$, assume the coefficient of technological innovation returns $\alpha = 30$ and the technological content of technological innovation $W = 18$. Fig.5a and fig.5b show the evolutionary path of the relationship between A and B under the conditions. In this case, the system has only one stable point (0,0), which means after a period of evolution, A and B can only adopt cooperation eventually to achieve bigger development, no matter what the initial ratios of cooperative innovation in A and B are.

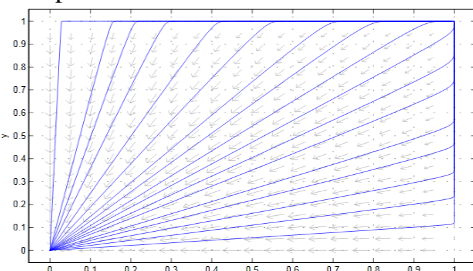


FIGURE 5a Phase Diagram in case 2

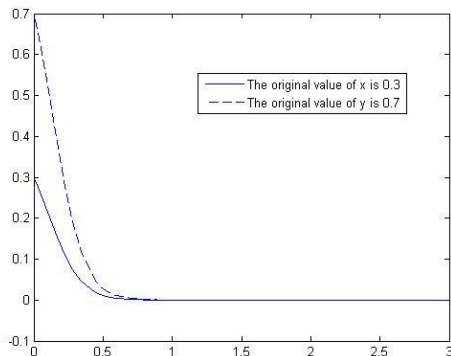


FIGURE 5b Evolutionary Path with stable point (0,0)

Values of parameters W , α and z can be adjusted respectively to analyse their effects on the game strategies of A and B.

(1) Change the value of W . Make $W = 18, 22, 26, 30, 35$ in turn. Fig.6a and fig.6b show that enterprise groups A and B fast approaches stable point (0,0), which represents cooperation as technological content increases.

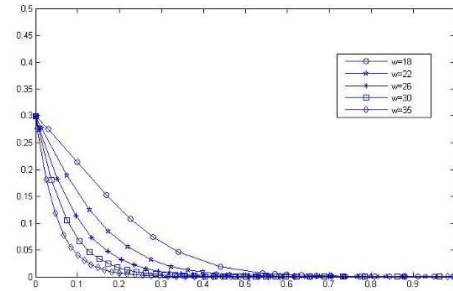


FIGURE 6a A's evolutionary Path figure as W increase

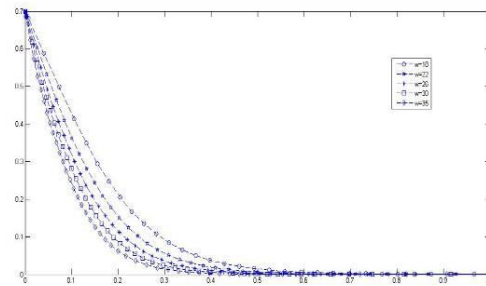


FIGURE 6b B's evolutionary Path figure as W increase

(2) Change the value of α . Make $\alpha = 30, 35, 40, 45, 50$ in turn. Fig.7a and fig.7b show the technological innovation revenues of A and B increase as the value of innovation coefficient α increases. This result promotes enterprises to adopt cooperative innovation.

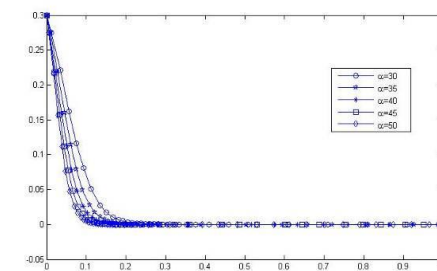


FIGURE 7a A's evolutionary Path figure as α increase

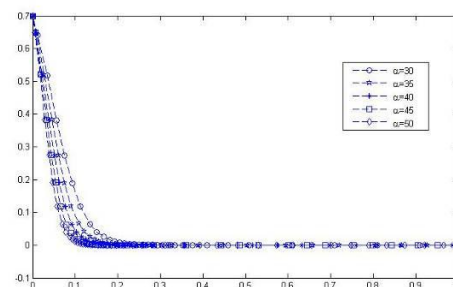


FIGURE 7b B's evolutionary Path figure as α increase

(3) Change the value of z . Make $z=0.4, 0.5, 0.6, 0.7, 0.8$ in turn. Fig. 8a and fig. 8b show that as z increases A and B both fast tend towards stable point $(0,0)$, which represents cooperation.

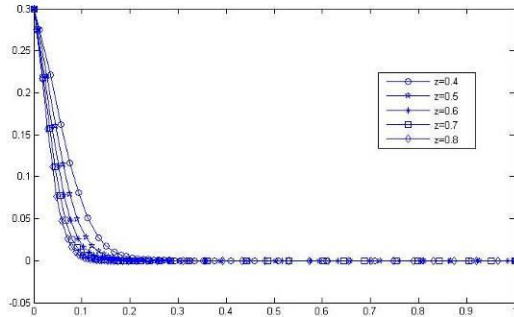


FIGURE 8a A's evolutionary Path figure as z increase

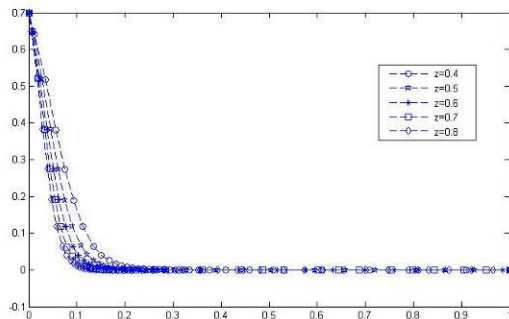


FIGURE 8b B's evolutionary Path figure as z increase

Case 3: To meet conditions of $D-F > 0$ and $G-E < 0$, assume the coefficient of technological innovation revenue $\alpha = 4$ and the technological content of technological innovation $W = 6$. Fig.9a and fig.9b show the evolutionary diagram and path in this condition. It can be seen the system has one stable point $(1,1)$, which means after a period of evolution, A and B eventually adopt noncooperation, no matter what the initial ratios of noncooperation in A and B are.

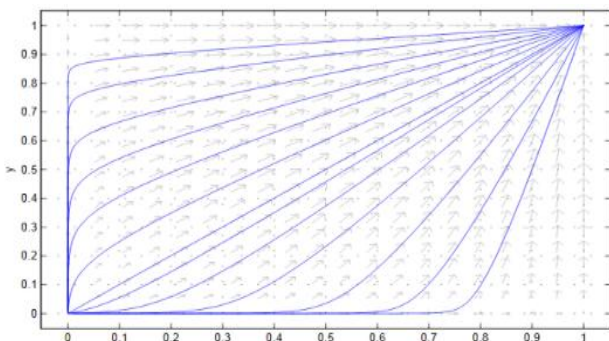


FIGURE 9a Phase Diagram in case 3

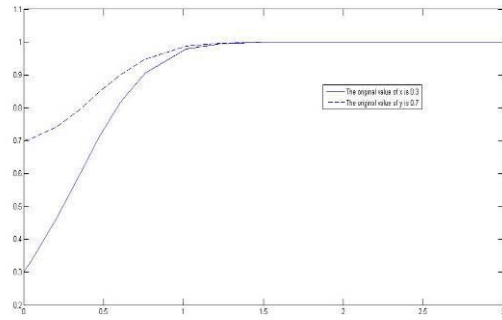


FIGURE 9b Evolutionary Path with stable point $(1,1)$

Values of parameters W , α and z can be adjusted respectively to analyse their effects on the game strategies of A and B.

(1) Change the value of the parameter of technological content W . Make $W=6, 12, 15, 18, 21$ in turn. Fig.10a and fig.10b show the evolutionary path of the relationship between A and B. Comparing fig.10a with fig.10b, we can see A slows down when approaching "1", while B is not sensitive to the increase of W . Therefore, as technological content increase, the evolution approaches point $(0,1)$ which means one party cooperates while the other party doesn't. It indicates the game result may have a qualitative change over time

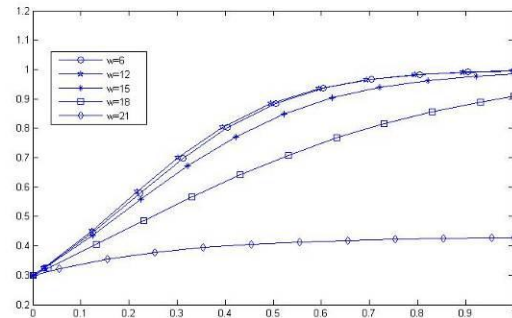


FIGURE 10a A's evolutionary Path figure as W increase

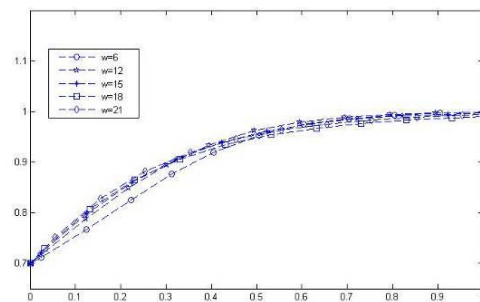


FIGURE 10b B's evolutionary Path figure as W increase

(2) Change the value of α . Make $\alpha = 4, 5, 6, 8, 10$ in turn. Fig.11a and fig.11b show that as technological innovation revenue (coefficient) increase, the system slows down when approaching stable point $(1,1)$.

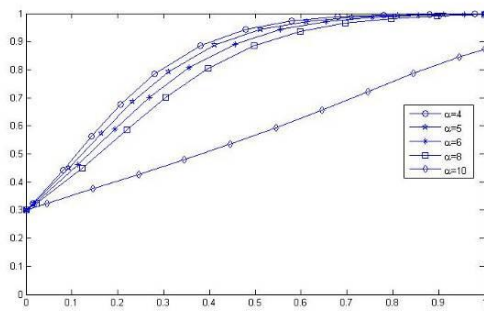


FIGURE 11a A's evolutionary Path figure as α increase

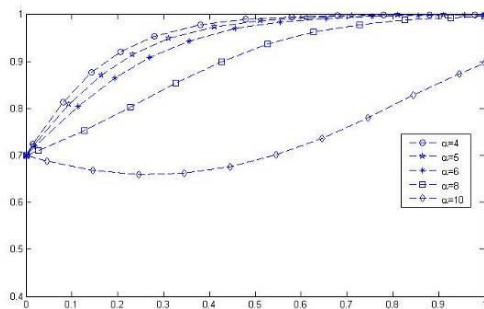


FIGURE 11b B's evolutionary Path figure as α increase

(3) Change the value of z . Make $z=0.4, 0.5, 0.6, 0.7, 0.8$ in turn. Fig.12a and fig.12b show that as the number of neutral enterprise increases, the system slows down when approaching stable point (1, 1).

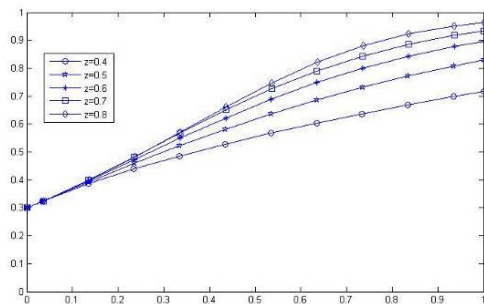


FIGURE 12a A's evolutionary Path figure as z increase

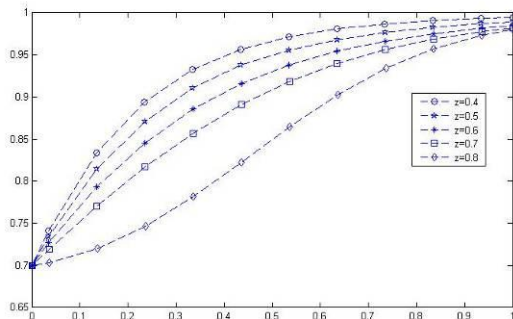


FIGURE 12b B's evolutionary Path figure as z increase

Case 4: To meet conditions of $D-F < 0$ and $G-E < 0$, assume the coefficient of technological

innovation revenue $\alpha = 8$ and the technological content of technological innovation $W = 25$. Fig.13a and fig.13b show the evolutionary diagram and path under this condition. According to these figures, A and B become stable at points (0,1) and (1,0). It means one party adopts cooperation while the other party adopts noncooperation. The final stable point is affected by the initial numbers of enterprise A and B adopting cooperation or noncooperation.

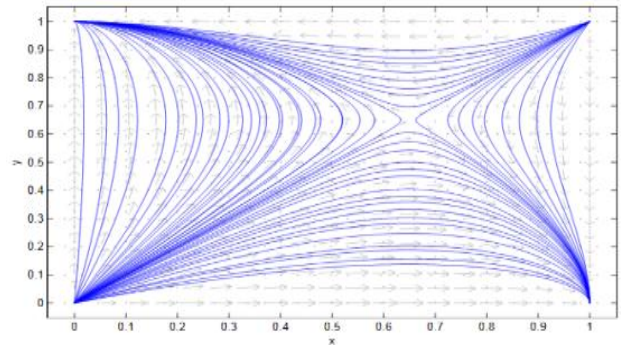


FIGURE 13a Phase Diagram in case 4

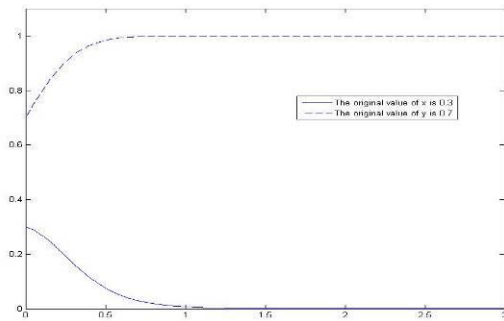


FIGURE 13b Evolutionary Path with stable point (0,1)

Values of parameters W , α and z can be adjusted respectively to analyze their effects on their game strategies of A and B.

(1) Change the value of W . Make $W = 25, 30, 35, 40, 45$ in turn. Fig.14a and fig.14b show A and B approach stable point (0, 1) faster as the value of W increases.

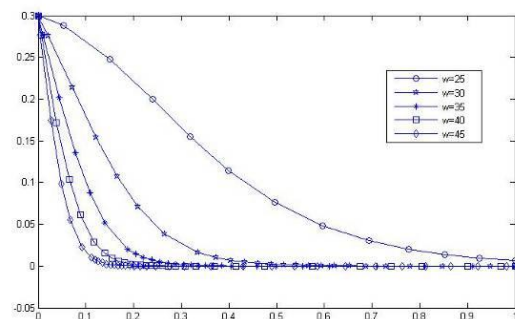


FIGURE 14a A's evolutionary Path figure as W increase

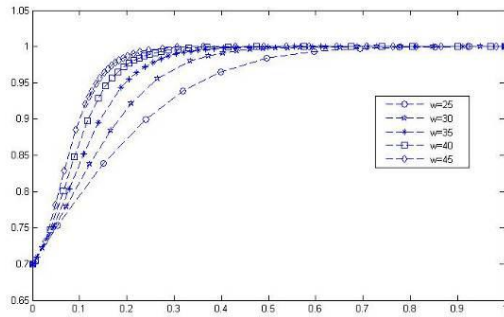


FIGURE 14b B's evolutionary Path figure as W increase

(2) Change the value of α . Make $\alpha = 8, 9, 12, 16, 20$ in turn. Fig.15a and fig.15b show that as the coefficient of technological innovation revenue increases, enterprise group A approaches the stable value "0" fast, while enterprise group B slows down when approaching the stable value point "1". It indicates as the technological innovation revenue (coefficient) increases, it's harder to approach the point (0,1), which means one party cooperates while the other party does not.

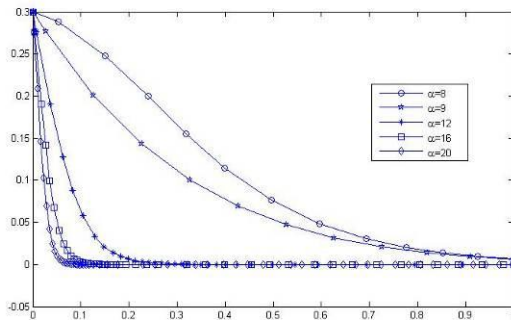


FIGURE 15a A's evolutionary Path figure as α increase

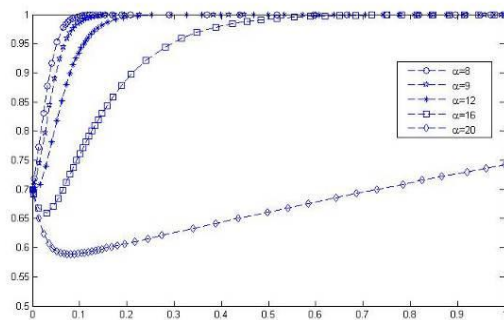


FIGURE 15b B's evolutionary Path figure as α increase

(3) Change the value of z . Make $z = 0.4, 0.5, 0.6, 0.7, 0.8$ in turn. Fig.16a and fig.16b show that enterprise group A approaches value point "0" faster, while enterprise group B slows down when approaching stable value point "1". It indicates that as the ratio of neutral enterprise increases, it's harder for the result of evolutionary game to approach point (0,1).

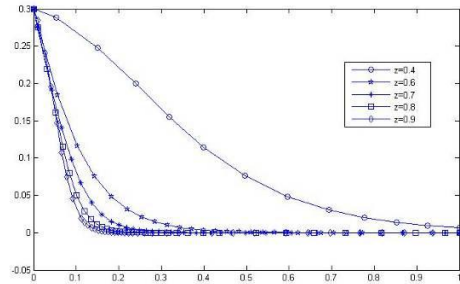


FIGURE 16a A's evolutionary Path figure as z increase

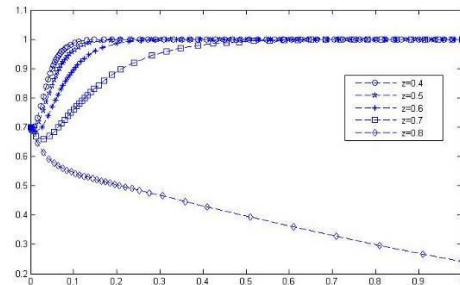


FIGURE 16b B's evolutionary Path figure as z increase

3.3 FURTHER DISCUSSION

According to the analysis on the four cases above, a further discussion is made in this section.

(1) The technological content of innovation object has great effects on inter-enterprise relationship. The discussion above shows that in case 1 and case 2, the increase in technological content makes the game result change into (cooperation, cooperation) eventually. The main reason for the result is, when carrying out an innovation activity with a high technological content, the enterprise tends to accomplish the innovation task with the help of other enterprises due to its limited innovation ability, and thus is willing to adopt cooperation. In case 3 and case 4, the increase in technological content makes two parties in the game fast approach a situation that one party cooperates while the other party doesn't. The reason is some enterprises begin to change innovation strategies facing the innovation risks brought by high technological content, and hope to accomplish the technological innovation task through the cooperation with others, as the technological content increases in case 3; in case 4, the enterprise of cooperation hopes to make cooperation more eagerly, while the enterprise of noncooperation is more reluctant to make cooperation, because of the increase in technological content. The noncooperation comes from two motives: first, facing the projects with high technological contents, some enterprise hope to steal the core technology from their partners through opportunistic behaviours; second, some enterprises doubt about the success rate of cooperative innovation, thus adopting neutrality (unwilling to cooperate) to avoid innovation risks. Over time, the results in case 3 and case 4 may both result in a situation that enterprises don't want to carry out innovation activities with high

technological contents any more, which is the last thing people want.

(2) Innovation revenue is the driving force for the establishment or maintenance of inter-enterprise relationship in technological innovations. As everyone knows, pursuing the maximization of benefit is the start point and one of end points of enterprise activities, so the analysis on technological innovation revenue has great effects on enterprises' innovation strategy selection. It can bring about a cooperation relationship, help to maintain cooperation, slow down the forming of noncooperation, and even bring about new cooperation relationships. It has been expressed clearly in the discussion above.

(3) Establishing and maintaining a certain amount of neutral relationships are very important for the cooperative innovation of enterprises. Different from general studies on competition and/or cooperation relationship, the paper subdivides the inter-enterprise technological innovation strategies into cooperation, neutrality and noncooperation. The study in the paper indicates the increase of the enterprises of neutrality eventually helps to form or approach a cooperation relationship of enterprises, no matter in which case. Therefore, identifying and maintaining neutral enterprises and keeping them from becoming competitors are very crucial in enterprise relationship management, which can lay a solid foundation for inter-enterprise cooperation in the future. Although the implementation of the strategy

needs certain cost, the cost is nothing compared with the cost of changing a competitor into a partner.

4 Conclusions

The paper classifies enterprises' technological innovation strategies as cooperation, noncooperation and neutrality using the idea and method of system analysis, and analyses the evolution of inter-enterprise relationship and strategy selection in technological innovation using the evolutionary game method and MATLAB simulation technique, and finally draws some valuable conclusions. It should be noticed that the study in the paper is still limited. For instance, for simplicity, the paper expresses the revenues of two parties both as $q_2\alpha W$ in the case that one party adopts competition while the other party adopts cooperation. In fact, the assumption may be different from the practical situation. Such problems should be solved in further and in-depth studies.

Acknowledgments

We acknowledge support from the National Natural Science Foundation of China under grant 71171135 and 71262011, Scientific Research Foundation of Jiangxi Provincial Education Department under grant GJJ13332 and Personnel Foundation of Jiangxi Province under grant 20122BCB23009

References

- [1] Li Ling 2006 The effects of information technology implementation on supply chain collaboration *International Journal of Internet and Enterprise Management* 4(2) 118-34
- [2] Xu S, Da Xu L 2011 *Information Technology and Management* 12(2) 51-4
- [3] Yi Yuyin, Xiao Tiaojun, Sheng Zhaohan 2005 *Journal of Shanghai Jiaotong University* 8(4) 80-7 (in Chinese)
- [4] Kale P, Singh H, Perlmutter H 2000 *Strategic Management Journal* 21(3) 217-37
- [5] Liu Zhaode, Zhang Weiguo 2008 *Science and Technology Management Research* 28(3) 20-22 (in Chinese)
- [6] Quintana-Garcia C, Benavides-Velasco C A 2004 *Technovation* 24(12) 927-38
- [7] Xu Liang, Zhang Zongyi, LongYong, Zhou Xu 2009 *Science Research Management* (1) 87-96. (in Chinese)
- [8] Liu Dehai 2008 *The Evolutionary Game Theory Analysis in Chinese Rural Labor Force Transfer* Metallurgical Industry Press: Beijing (in Chinese)
- [9] Zhien Ma, Yicang Zhou 2012 *Ordinary Differential Equation Qualitative and Stability Analysis* Science Press: Beijing (in Chinese)

Authors



Nannan Duan, born in January, 1981, Nanchang, Jiangxi, China

Current position, grades: School of Economics and Management, East China JiaoTong University.

University studies: doctoral candidate of Business School, the University of Shanghai for Science and Technology.

Scientific interest: Technological Innovation, Supply and Demand Network.

Publications: more than 10 papers published in various journals.



Fuyuan Xu, born in September, 1948, Shanghai, China

Current position, grades: professor, doctoral supervisor, head of the Management Science and Engineering Doctoral Programme, the University of Shanghai for Science and Technology.

University studies: Graduated from University of Lumiere-Lyon II in 1984, received a doctor's degree in Computer Science.

Scientific interest: Supply and Demand Network.

Publications: more than 130 papers published in various journals.

Experience: more than 30 scientific research projects, more than 20 teaching materials.



Ming Ni, born in February, 1974, Nanchang, China

Current position, grades: assistant professor, assistant dean of School of Economics and Management, East China JiaoTong University

University studies: Graduated from University of Shanghai for Science and Technology in 2006, doctor's degree in management.

Scientific interest: Supply and Demand Network.

Publications: more than 30 papers published in various journals.

Experience: more than 15 scientific research projects, more than 10 teaching materials.

Subpixel edge extraction of part ant colony optimization-based and dimensional measurement

Lixiong Gong^{1*}, Xiangsheng Kong², Yong Liu³, Min Huang⁴

¹*School of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054*

²*Department of Computer and Information Engineering, Xinxiang University, Xinxiang He'nan 453003*

³*School of Construction Machinery, Chang'an University, Xi'an Shanxi 710064*

⁴*MBA Education Center, Chongqing University of Technology, Chongqing 400050*

Received 12 May 2014, www.tsi.lv

Abstract

Put forward a method combined improved ant colony and Zernike moment to detect image subpixel edge aiming at traditional ant colony algorithm's drawback of long time consumption and easily to be affected by noise. The methods improved parameters from clustering centre setting, clustering operator selecting and pheromone updating, then extracted subpixel image edge based on Zernike moments. Therefore, the result of image edge extraction is good and effective. Lastly, least square fitting is used to locate coordination of image edge and bearing of SKF 32308 J2/Q dimensions such as inner and outer diameters were measured. The result shows that the algorithm proposed can well to measure circular parts dimensions and has high precision.

Keywords: Subpixel, Ant Colony Optimization, Bearing, edge detection

1 Introduction

Vision inspection as non-contact measurement has widely been paid a wide attention, and was focused on geometry and surface quality detection [1, 2]. Method of vision inspection can be divided into visual detection region-based and edge detection according to different principle of computer vision [3, 4]. Edge detection depends on the change of part image grayscale, and is high stable and fixed. Therefore, edge extraction algorithm with high accuracy is a basis of vision inspection. Commonly algorithms of edge inspection locate image edge by operator depending upon change of image neighbourhood grey, e.g., Roberts, Prewitt, Sobel, Laplacian and Canny operators. It is difficult to satisfied with result of edge detection for all those traditional operators because of large calculation, a long time, poor anti-interference ability [5, 6]. Moreover, pixel-level edge detection has been unable to meet the need of detection with the continuous increasing of machining precision. Many scholars began to study subpixel image edge extraction algorithms. e.g., Lyvers et al proposed spatial moment method for edge detection, but calculation are huge [7]. Nowadays, ant colony algorithm was applied to digital image processing by experts all over the world. e.g., Hoseine studied hybrid ant colony algorithm to optimize image in details, effectively enhance the contrast of the image [8]. However, the noise cannot be effectively removed due to the same selection probability between noise and edge during ant searching process. XUE Qin

presented image segmentation algorithm based on ant colony gradient operator, which it can be segmented from the similar background and target grey value image [9]. WANG Dong applied Sobel operator and ant colony algorithm to search local edge, which detect more effective infrared image [10]. However, above improved algorithms are unable to avoid drawbacks of the gradient and ant colony algorithm that time for searching non-edge region is long, and detection of edge orientation is limited. Therefore, a method combined improved ant colony and Zernike moment to detect image subpixel edge was proposed, the method improved parameters from clustering centre setting, clustering operator selecting and pheromone updating, and then extracted subpixel image edge with high speed and accuracy based on Zernike moments. The dimension of type SKF 32308 J2/Q bearing was measured in the experiment to testify the accuracy of measurement.

2 Ant Colony Optimization

2.1 ANT COLONY ALGORITHM

Ant colony algorithm (ACA) is an optimization algorithm proposed by Italian researchers M. Dorigo etc. in 1991. In the past ten years, ACA has been successfully applied to the combinatorial optimization problems [11-13]. Feeding on ants, they can always find the optimal path from nest to food source. When the path is blocked, ants can bypass the barrier and find other optimal path again.

* *Corresponding author* e-mail: herogong2001@sohu.com

Many ants foraging behaviour made up a positive feedback of information process that formed the optimal path. Its essence is a constant clustering process. The Formulation of ACA in image edge detection can be described as follows.

Each pixel $X_j(j=1,2,\dots,N)$ from original image is supposed to an ant, and then the pixel is taken as two-dimensional vectors of grey and gradient. Now the similarity between pixel X_i and X_j is represented by Euclidean distance, so

$$d_{ij} = \frac{1}{\eta_{ij}} = \sqrt{\sum_{k=1}^m p_k (x_{ik} - x_{jk})^2}, \quad (1)$$

where m is ant pixel dimension and m equal to 2, p represents weight factor, η_{ij} is heuristic function. The formula of ant pheromones in paths is computed as follows.

$$\tau_{ij}(t) = \begin{cases} 0 & (d_{ij} \leq r) \\ 1 & (d_{ij} > r) \end{cases}, \quad (2)$$

where r is given threshold. When distance of ant pixel d_{ij} is less than or equal to r , ant pheromones of the path is 1, or else is 0. Any pixel transition probability from X_i to X_j is described that

$$P_{ij}(t) = \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)}{\sum_{s \in S} \tau_{is}^\alpha(t)\eta_{is}^\beta(t)}, \quad (3)$$

where α and β are selected path impact factors of ant clustering, respectively. S is feasible path sets of ant, and s is an optional path. Pheromone of each path is adjusted by formula (4) with the continuous ant's moving.

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \tau_{ij}, \quad (4)$$

where ρ represents volatile coefficient of pheromone, $1-\rho$ is residual factor of pheromone, the scope of ρ belongs 0 to 1. τ_{ij} is incremental pheromone in loop, and computed by formula (5).

$$\tau_{ij} = \sum_{k=1}^m \tau_{ij}^k. \quad (5)$$

And then

$$\Delta \tau_{ij}^k = \frac{Q}{d_{ij}}, \quad (6)$$

where Q is an integer greater than 0.

2.2 OPTIMIZED ANT COLONY ALGORITHM

The time of ant search process is very long because of blind and random finding. Therefore, the paper sets clustering centre to guide ant to run together, and proposes pheromone updated algorithm to void premature, which improve accuracy and rapidity of edge detection.

(1) Initial clustering centre setting

There are some image segmentation errors because one-dimensional grey histogram used by the traditional ant colony algorithm cannot express local spatial information. The two-dimensional grey histogram is clustered in the paper, namely the initial cluster centre is composed of histogram of the pixel and neighbourhood mean grey distribution. The steps are as follows.

Step1. Draw 2D grayscale of measured original image.

Step2. For the best one-dimensional projection on two-dimensional grayscale, reduce dimension.

Step3. Select peak as the initial clustering centre from one-dimensional projection histogram.

(2) Classification and calculation for clustering centre

Taking the pixel grey value, neighbourhood characteristic value, the pixel point gradient value as the first feature point, second point, third point, and when the gradient value of clustering centre is zero, then setting neighbourhood features as eight. If there are large points of clustering centre relatively, we can decide the clustering centre into image edge point centre, then set neighbourhood feature value as six, else take it as noise points clustering centre, set neighbourhood feature value as 3. The distance from pixel points to clustering centre is that:

$$d_{ij} = \sqrt{\sum_{k=1}^3 p_k (x_{ik} - c_{jk})^2}, \quad (7)$$

where C_j is clustering centre, and $C_j = (c_{j1}, c_{j2}, \dots, c_{jk}), j=1, 2, \dots, M$.

(3) Updating pheromone

Traditional ant colony algorithm identified similar typical ants to the same class, which can lead to pheromone that is not searched reduce to zero and perhaps emerge wrong image segmentation problems. Therefore, this paper proposed global adaptive updated pheromone algorithm.

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \tau, \quad (8)$$

$$\tau_{mean} = mean(\tau_{i1}, \tau_{i2}, \dots, \tau_{is}), \quad (9)$$

$$\begin{cases} \tau_{ij} \geq \tau_{mean} & \tau = \tau_{mean} \\ \tau_{ij} < \tau_{mean} & \tau = \tau_{ij} \end{cases}, \quad (10)$$

where $\tau_{i1}, \tau_{i2}, \dots, \tau_{is}$ is a respective pheromone increment optional path, and τ_{mean} is mean of pheromone increment. It can be seen pheromone increment is mean value when a path pheromone increment is exceed to mean value, which can ensure ants search are not in less paths. The optimized ant colony algorithm is described as follows.

Step1. Initialize ant colony parameters and setting number of cycles.

Step2. Use the two-dimensional grayscale measured image to find clustering centre and partitioning image edge, target points, noise and others by Laplace operator.

Step3. Compute similar function, heuristic function between pixel points and clustering centre of measured image.

Step4. Select the maximum probability of clustering centre.

Step5. Update pheromone and calculating the clustering centre value.

Step6. Execute Step 4 again, and moving in circles, until it reaches the set of cycle number.

Subpixel edge detection with Zernike moment

It is common algorithm to extract subpixel image edge using spatial moment. Reference [14] computed four parameters according to six spatial moments. However the algorithm is very complex and inefficient. In this paper, we proposed Zernike moment to detect image edge that reduce computation and complexity, the algorithm can calculate four parameters only by three Zernike moments.

Definition1. The n^{th} order Zernike moment of image $f(x, y)$ is

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(\rho, \theta) dx dy, \quad (11)$$

where $V_{nm}^*(\rho, \theta)$ represents integral function, m and n are integers, which are satisfied with condition that $n \geq 0$ and $n - |m|$ is even number.

Figure1 is ideal unit circle step-edge model, where k is step-gradients, h is grey background and the grey value of shade is $h+k$, and l is vertical distance from circle centre to edge, ϕ is angle between vertical edge and X axis. If image is rotated angle ϕ clockwise, which is parallel the Y axis, as shown in Figure 2, and Figure3 is profile of rotated ideal step-edge model.

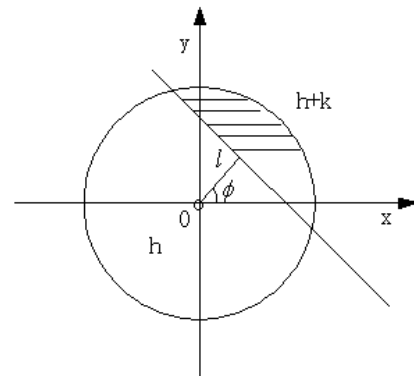


FIGURE 1 Ideal step-edge model

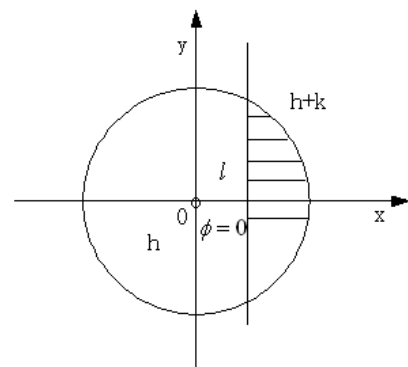


FIGURE 2 Rotated ideal step-edge model

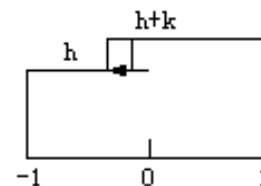


FIGURE 3 Profile of rotated ideal step-edge

Setting rotated image $f'(x, y)$ is symmetric about X axis, then

$$\iint_{x^2+y^2 \leq 1} f'(x, y) y dy dx = 0. \quad (12)$$

The important feature is rotational invariant of Zernike moment. If an image is rotated by angle ϕ , and only the phase is shifted before and after rotation, the amplitude keeps unchanged. The step Zernike moments of rotational image can be derived combined with Figure 3.

$$A'_{00} = A_{00} = h\pi + \frac{k}{2} \pi - k \sin^{-1}(l) - kl\sqrt{1-l^2}, \quad (13)$$

$$A'_{11} = A_{11}e^{j\phi} = \frac{2k}{3}(1-l^2)^{2/3}, \tag{14}$$

$$A'_{20} = A_{11} = \frac{2k}{3}l(1-l^2)^{2/3}. \tag{15}$$

Simultaneous equation, then

$$l = \frac{A'_{20}}{A'_{11}} \quad K = \frac{3A'_{11}}{2(1-l^2)^{3/2}}. \tag{16}$$

Template size of edge detection is $N \times N$, the calculation of Zernike moments of edge is shown in formula (17).

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{Nl}{2} \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}, \tag{17}$$

where (x, y) is origin of coordinates in Figure 1, (x_s, y_s) is subpixel coordinate of image edge. In the algorithm, four parameters can be computed by three Zernike moments, i.e., ϕ, l, k, h . And $\delta = 1/\sqrt{2}$ pixel, τ is threshold that can be chosen freely in fact. Now algorithm of subpixel edge detection Zernike moments-based is described as follows.

Step1. Calculate moments A_{00} , A_{11} and A_{20} using image template an convolution.

Step2. Calculate moments A_{00} , A_{11} and A_{20} after rotated ϕ angle by means of rotational invariant.

Step3. Calculate l and k by formula (16).

Step4. Determine step-grey threshold of τ .

Step5. Find pixel points that meet condition of $k \geq \tau \cap l \leq \delta$, then the point is subpixel edge, we can compute the subpixel coordinate by formula (17).

3 Experiment and Discussion

In experiment, bearing of SKF 32308 J2/Q was studied and tested. Geometric parameters of bearing mainly include inner and outer diameters. The camera used is Costar brand, and the lens is a Pentax made in Japan. The original image is obtained as shown in Figure 4.

The objective of this experiment is to acquire bearing image, measure geometric size and verify the accuracy of image edge detection proposed algorithm by comparison. The steps and image-processing algorithm are presented as follows.



FIGURE 4 SKF 32308 J2/Q bearing original image

Step 1. Bear image pre-processing and enhancing contrast

The pre-processing of bearing image is designed by nonlinear spatial filter, which the centre is -8 and other is Laplace filters that mean value is one with 3×3 to enhance image edge and deduce noise, as shown in Figure5. Figure6 is enhanced grey image after pre-processing. It can be seen from Figure6 that bearing image can be described in detail and help to divide image edge based on the ant colony algorithm in the next step.



FIGURE 5 Bearing image after spatial filter

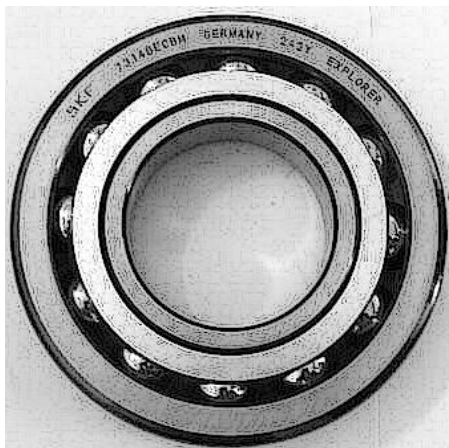


FIGURE 6 Bearing image enhanced processing

Step 2. Image segmentation optimized ant colony algorithm-based aiming at pre-processing image

In the step, the image edge of Figure6 is detected using optimized ant colony algorithm. In the experiment, we compare different image edge segement results for Sobel operator, Canny operator, traditional ant colony algorithm and optimized ant colony algorithm proposed in the paper. The experimental results are shown in Figure7, Figure8, Figure9 and Figure10.



FIGURE 7 Sobel operator image sementation

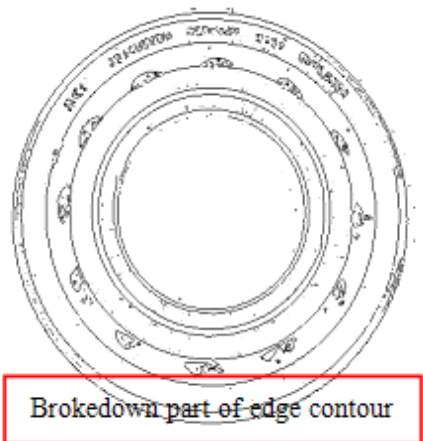


FIGURE 8 Canny operator image sementation

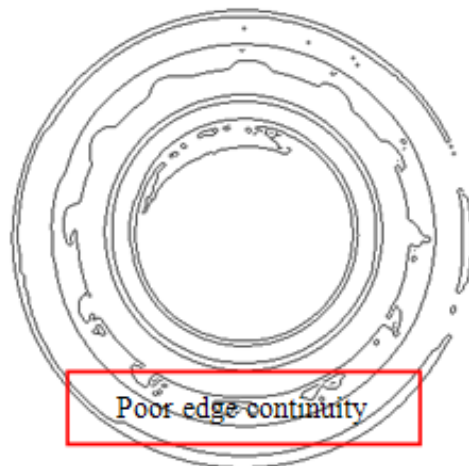


FIGURE 9 Traditional ant colony algorithm



FIGURE 10Optimized ant colony algorithm

From above Figures, it can be known Sobel operator are poor results in unclear edge of image segmentation, and Canny operator have some broke down in edge contour, traditional ant colony algorithm appear poor edge continuity and easy to affected by noise. Obviously, optimized ant colony algorithm proposed in the paper has intact and full edge contours of image, and speed of segmentation is rapid and effective.

Step 3. Image subpixel edge detection Zernike moment-based

Assuming each pixel points gray value $f(x, y)$ is constant, then the gray of neighbourhood of image (x, y) and M of template $9*9$ can be calculated by convolution, we can get Zernike moment and edge parameters. The subpixel edge points of bearing inner and outer rings can be determined according to judgement criterion of edge.

Step 4. The least square fitting and determining the inner and outer diameters of bearing and circular centre

While the bearing subpixel edge is located, the edge points must be fitted in order to define the bearing inner, outer diameter and circular centre of geometry parameters. The least square fitting is very effective in all of sub-pixel fitting methods; reference [15] explained the

least square algorithm. Figure 11 draws bearing inner and outer diameters and circular centre location.

Step 5. Calibrating bearing and converting pixel to real dimension (unit: mm).

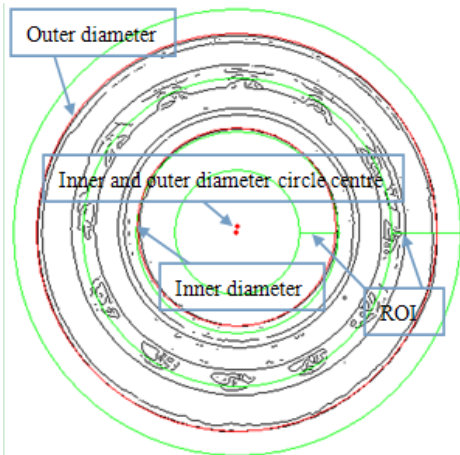


FIGURE 11 Subpixel fitting of bearing inner and outer diameters and circular centre

TABLE 1 Measured results

Number	Coordinate of inner diameter circular centre (pixel)	Inner diameter dimension (mm)	Coordinate of outer diameter circular centre (pixel)	Outer diameter dimension (mm)
1	(304.01, 245.32)	40.01	(303.90, 248.94)	90.02
2	(303.88, 248.92)	39.99	(304.01, 245.33)	89.98
3	(304.16, 248.24)	40.00	(303.52, 245.23)	89.97
4	(302.94, 246.96)	40.00	(304.12, 245.23)	90.01
5	(304.89, 247.74)	40.02	(303.13, 246.11)	90.03
6	(304.20, 245.12)	39.97	(304.22, 245.83)	90.02
7	(304.22, 245.41)	40.01	(304.45, 245.43)	89.99
8	(305.44, 245.31)	39.97	(305.09, 246.31)	90.02
9	(301.36, 248.21)	39.97	(304.12, 246.23)	89.98
10	(303.17, 249.28)	40.02	(305.12, 246.35)	90.02

The inner and outer diameters of standard bearing SKF 32308 J2/Q are respective 40mm and 90mm. It can be seen from Table 1 that mean error of measured inner diameter is 0.016mm, and outer diameter is 0.020mm, so the measurement accuracy meets measured demands.

4 Conclusions

Image edge detection is an important field of image processing. A method combined improved ant colony and Zernike moment is proposed to detect image subpixel edge in the paper. There are some conclusions are drawn in the paper.

(1) Ant colony algorithm is wide application for image classification and edge detection, but the precision is limited and time cost is large. Methods proposed in the paper optimize parameters from clustering centre setting, clustering operator selecting and pheromone updating. The result of image edge extraction is good and effective.

Selecting standard bearing to calibrate and convert pixels size to real dimension by formula (18).

$$K = d/d_{pixel} , \tag{18}$$

where *K* is converted equivalence by Calibration, the value is 0.5455 pixels/mm in the paper.

After above steps are completed, we select 10 bearings that type is SKF 32308 J2/Q to measure the inner and outer diameters. Measured results were shown in Table 1.

(2) The image subpixel edge can quickly be identified by Zernike moments, which only need three Zernike moments to calculate four parameters. Due to reduce computation and complexity, the algorithm can be applied to image subpixel identification online.


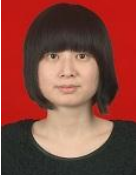
(3) Compared to standard bearing of SKF 32308 J2/Q, the parameter errors measured inner and outer diameters are within allowed scope, and meet the requirements of precision measurement.

Acknowledgments

This study was jointly supported by Chongqing Basic and Advanced Research Project (No. cstc2013jcyjA60002), Development Fund of Chongqing University of Technology (No. SK201307) and Opening Program of Management Science and Engineering of Chongqing University of Technology.

References

- [1] Malassiotis S, Strytzis M G 2003 Stereo vision system for precision dimensional inspection of 3D Holes *Machine Vision and Applications* **15**(2) 101-13
- [2] Hong ZH Y, Yang Y Y, Deng SH X 2010 Research on laser measurement system of hot-rolling wheel size *Chinese Journal of Electronic Measurement and Instrument* **24**(7) 686-91 (In Chinese)
- [3] Sun Qiucheng, Tan Qingchang, Tang Wusheng, et al. 2010 Shaft diameter measuring method based on CCD *Journal of Jilin University: Engineering and Technology Edition* **40**(5) 1273-7 (In Chinese)
- [4] Dong Hongyan 2008 *Research on some techniques in edge detection* Changsha: National University of Defense Technology
- [5] Canny J 1986, A computational approach to edge detection *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6) 679-98
- [6] Zhang Yu-jin 2001 *Image segmentation* Science Press:Beijing
- [7] Lyvers E P, Mitchell O R, Akey M L, et al. 1989 Subpixel Measurements using a moment-based edge operator *IEEE Transactions on Pattern Analysis and Machine Intelligence* (11) 1293-309
- [8] Pourya Hoseini, Mahrokh G, Shayesteh 2013 Efficient contrast enhancement of images using hybrid ant colony optimisation, genetic algorithm, and simulated annealing *Digital Signal Processing* **23** 879-93
- [9] Xue Qin, Chen Wei, Luo Jun-qi 2007 Research on image segmentation by ant colony algorithm based on gratitude operator *Computer Engineering and Design* **28**(23) 5660-3
- [10] Wang Dong, Zhang Jing-Zhou 2011 Edge Detection of Infrared Image Based on Sobel and Ant Colony Algorithm *Journal of Henan University of Science and Technology: Natural Science* **32**(6) 38-41
- [11] Dorigo M, Birattari M, Stutzle T 2006 Ant colony optimization: artificial ants as a computational intelligence technique *IEEE Computational Intelligence Magazine* **11** 28-39
- [12] Dorigo M, Maniezzo V, Colomi A 1996 Man, and Cybernetics, Ant System: optimization by a colony of cooperating agents *IEEE Trans on System* **26**(1) 29-41
- [13] Zhang Wei, Pan Xiao-hong, Liu Zhi, et al. 2012 Manufacturing service scheduling strategy on cloud model ant colony optimization *Computer Integrated Manufacturing Systems* **18**(1) 201-7
- [14] Yu H H 1997 A visual search system for video and image databases *Proc of International Conference on Multimedia Computing and Systems, IEEE Computer Society: Washinton DC* 517-24
- [15] Sun Wei, Zhang Xiao-rui, Tang Hui-qiang, et al. 2011 Lane Coordination Detection Based on Hough Transformation and Least Square Fitting *Opto-Electronic Engineering* **38**(10) 13-9

Authors	
	<p>Lixiong Gong</p> <p>Current positions, grade: Chongqing University of Technology, Lecturer. University studies: Wuhan University of Technology (2007-2010), Ph.D. Scientific interest: Quality control and measurement, Vision image detection, Advanced manufacturing and informatization. Publications: more than 20 publications</p>
	<p>Xiangsheng Kong</p> <p>Current positions, grade: Xinxiang University, Lecturer. University studies: PLA Information Engineering University, China (1999-2003), Bachelor. Scientific interest: system analysis & design, software testing. Publications: 9 publications.</p>
	<p>Yong Liu</p> <p>Current positions, grade: Doctoral candidate of Chang'an University. Scientific interest: signal processing, Advanced manufacturing.</p>
	<p>Min Huang</p> <p>Current positions, grade: Master candidate of Chongqing University of Technology. Scientific interest: ERP, Management information system, MBA</p>

Application of PID-type iterative learning control for DC motor

Hong Li*

School of Information Engineering, Xianyang Normal University, Xianyang, Shaanxi, China, 712000

Received 1 March 2014, www.tsi.lv

Abstract

Iterative learning control is a new control technology, which is a branch of intelligent control theory and particularly suitable for the controlled object with repetitive motion. In this paper, a PID-type iterative learning control for DC motor based on the characteristics of repeating motion of DC motor was proposed and the convergence of iterative learning control algorithm was analysed. The input of controlled system in current cycle was amended by the error achieved between the system output and the desired trajectory in previous iteration. It was tested that PID-type ILC had good performance and stability through a large number of simulations and the experiments of the velocity tracking are done by MATLAB software. The results showed that the velocity tracking precision of DC motor was higher and the error was smaller with the increasing number of iterations. The velocity tracking error was close to zero. It was also shown that the motor could fully track the given desired trajectory in some certain iteration. It was also revealed from simulation results that the proposed control strategy was valid and effective for the DC motor.

Keywords: PID control, DC motor, PID-type iterative learning control

1 Introduction

With the feature of simple structure, small size, light weight, reliable operation, easy maintenance, good servo performance, fast response speed and good stability, the DC motor is widely applied in the fields of servo system, factory automation and defence industry [1]. In recent years, with the rapid developing of electronic technology and micro-controller technology, it has a much higher demand to the velocity control precision of DC motor when it executes certain task repetitively. Owing to the difficulty of DC motor in setting up accurate mathematical model, the selection of control method seems utmost important in order to allow the system to own higher control precision, stability and higher tracking performance.

In 1984, Arimoto et al. first introduced the initial explicit formulation of iterative learning control [2, 3]. Since the ILC algorithm was proposed, a very large number of approaches have been considered. Its control actually imitates human learning character, achieves tracking task in limited time of simple in principle and needs higher control performance. The input of controlled system in current cycle is amended by the error achieved between the system output and the desired trajectory in previous iteration. ILC is a technique to control the systems with a defined task repetitively in a limited and constant time interval. Ref [4] proposed the robust optimal design problem and analyzed the convergence of iterative learning control. A novel model-based method was presented so that ILC based on a quadratic performance criterion was revisited [5]. Ref [6] designed the ILC algorithms based on parametric optimization

approach. ILC is widely concerned because of no precise mathematical model [7-9]. Lee applied ILC technique to improve tracking control performance in batch processes [10]. Ref [11] offered a 2-Dimensional systems theory based ILC. A driven data constrained norm-optimal iterative learning control framework was proposed [12]. Generally speaking, iterative learning control system can realize precise control characteristic by self-regulation disvelocity without the precise mathematical model. Therefore, it promotes a great development of D-type iterative learning control [13], P-type iterative learning control [14, 15], PD-type iterative learning control [16], PI-type iterative learning control [17], PID-type iterative learning control theories [18-24]. In recent years, iterative learning control theory is rapid developed both at home and abroad, and there are many achievements.

With the continual advances of control theory, the PID controller is still the most commonly used controller in the process control industry [25]. Traditional PID control can't track the real-time velocity timely and precisely, and its control is poorer. Liu Haishan directly applied PID control for the control of brushless DC motor. This method has strong robustness and poor traceability [26]. Jia Hongping applied sliding mode variable structure control for the control of brushless DC motor, and it has strong robustness, good traceability and buffeting [27].

This paper applied PID-type iterative learning control algorithm for the control of DC motor based on repetitive operation DC motor system. The convergence of the algorithm is analysed in theory and simulated for the mathematical model of DC motor in MATLAB. The results show that the simulation result agrees with the

* *Corresponding author* e-mail: 110508@sina.com

theoretical analysis. With the increasing of iterations, the accuracy is higher and higher, the error is smaller and smaller which is close to zero. In addition, the motor can fully track the given desired trajectory in some certain iterations.

The paper is organized as follows. The mathematical model of the DC motor is established in the Section 2. In Section 3, the PID-Type iterative learning control algorithm is discussed and designed. In Section 4, we make numerical simulations make for the DC motor transmission system using the proposed control method. Conclusions are drawn in Section 5.

2 Mathematical Model

In modern industry, DC motor has been the execution terminal extensively used in servo system. This paper regarded brushless electromagnetic DC motor as research object and the mathematical model is set up for it [28]. Figure 1 is the working principle and equivalent circuit diagram of DC motor.

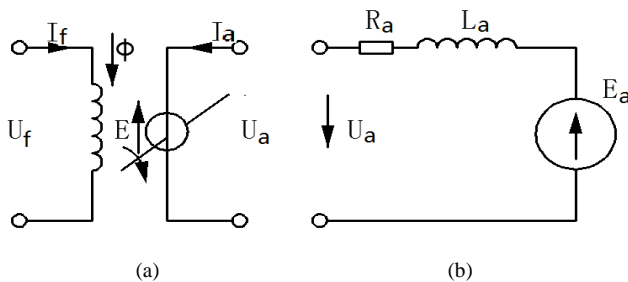


FIGURE 1 Diagram of working principle and the equivalent circuit of DC motor ((a) working principle, (b) equivalent circuit)

Armature circuit voltage balance equation can be obtained:

$$U_a = E_a + R_a I_a = C_E \Phi_n + I_a R_a = K_E n + R_a I_a \quad (1)$$

Equation (2) is dynamic equation of DC motor:

$$u_a = K_E n + R_a i_a + L_a \frac{di_a}{dt} \quad (2)$$

where R_a is the loop resistance. I_a is loop current. E_a is loop induced electromotive force. U_a is the terminal voltage of circuit. C_E is the constant of electromotive force. n is the motor speed. K_E is the electromotive force generated by unit speed. Φ_n is the flux.

Equation (3) is electrodynamic balance equation:

$$T_e = B\omega + T_L + J \frac{d\omega}{dt} = C_T \Phi i_a = K_T i_a \quad (3)$$

where T_e is instant electromagnetic torque. T_L is load

torque. B is damping coefficient. J is rotational inertia. K_T is a constant of torque. ω is angel velocity.

Supposing the initial conditions of the motor is zero, and the load of the motor is constant. The transfer function can be obtained after the LAPLACE conversion:

$$G(s) = \frac{\omega(s)}{U(s)} = \frac{K_T}{L_a J s^2 + (R_a J + L_a B) s + R_a B + K_T K_E} \quad (4)$$

3 PID-Type Iterative Learning Control Algorithm

Iterative learning control is mainly to learn in repetition, and improve certain control objective through iteration control. The algorithm is simple, and it can realize the actual motion track of the unknown object in the given time horizon. The algorithm does not rely on the mathematical model of the controlled plant.

The PID-type iterative learning control law of linear time-varying system was given [2]:

$$u_{k+1}(t) = u_k(t) + \Gamma \dot{e}_{k+1}(t) + \Phi e_{k+1}(t) + \Psi \int_0^t e_{k+1}(\tau) d\tau \quad (5)$$

The form of singular linear continuous system is:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (6)$$

In the formula, Γ , Φ , Ψ are gain matrixes. k is iterations. A , B , C are real matrixes.

If the systems, which are described in Equations (5) and 6 satisfy the following conditions:

$$(1) \|I - CB\Gamma\| \leq \bar{\rho} < 1;$$

(2) Supposing the initial condition of each iteration is the same, namely, $x_k(0) = x_0$, ($k=1,2,3,\dots$), $y_0(0) = y_d(0)$. There is $y_k(t) = y_D(t)$, $\forall t \in [0, T]$ when $k \rightarrow \infty$.

Proof:

In accordance with reference [13], it can be inferred from Equation (6) and condition (2): $y_{k+1}(0) = Cx_{k+1}(0) = Cx_k(0) = y_k(0)$.

Therefore, $e_k(0) = 0$, the output error of the time of $k+1$ is:

$$e_{k+1}(t) = e_k(t) - \int_0^t C\varphi(t, \tau) B(\tau) [\Gamma(\tau) \dot{e}_k(\tau) + L(\tau) e_k(\tau) + \Psi(\tau) \int_0^\tau e_k(\delta) d\delta] d\tau \quad (7)$$

Supposing $G(t, \tau) = C(t)B(\tau)\Gamma(\tau)$, then

$$\int_0^t G(t, \tau) \dot{e}_k(\tau) d\tau = G(t, \tau) e_k(\tau) \Big|_0^t - \int_0^t \frac{\partial}{\partial \tau} G(t, \tau) e_k(\tau) d\tau = CB(t) \Gamma(t) e_k(t) - \int_0^t \frac{\partial}{\partial \tau} G(t, \tau) e_k(\tau) d\tau \quad (8)$$

Substitute (8) into (7), then

$$e_{k+1}(t) = [I - CB(t) \Gamma(t)] e_k(t) + \int_0^t \frac{\partial}{\partial \tau} G(t, \tau) e_k(\tau) d\tau - \int_0^t C \varphi(t, \tau) B(\tau) L(\tau) e_k(\tau) d\tau - \int_0^t \int_0^\tau C \varphi(t, \tau) B(\tau) \Psi(\tau) e_k(\sigma) d\sigma d\tau \quad (9)$$

Make bound norm of both ends of (9), then

$$\begin{aligned} \|e_{k+1}(t)\| &\leq \|I - C(t)B(t)\Gamma(t)\| \|e_k(t)\| + \int_0^t \left\| \frac{\partial}{\partial \tau} G(t, \tau) \right\| \|e_k(\tau)\| d\tau + \int_0^t \|C\varphi(t, \tau)B(\tau)L(\tau)\| \|e_k(\tau)\| d\tau + \int_0^t \int_0^\tau \|C\varphi(t, \tau)B(\tau)\Psi(\tau)\| \|e_k(\sigma)\| d\sigma d\tau \\ &\|I - C(t)B(t)\Gamma(t)\| \|e_k(t)\| + \int_0^t b_1 \|e_k(\tau)\| d\tau + \int_0^t \int_0^\tau b_2 \|e_k(\sigma)\| d\sigma d\tau \end{aligned} \quad (10)$$

where $b_1 = \max \left\{ \sup_{t, \tau \in [0, T]} \left\| \frac{\partial}{\partial \tau} G(t, \tau) \right\|, \sup_{t, \tau \in [0, T]} \|C(t)\Phi(t, \tau)B(\tau)L(\tau)\| \right\}$, $b_2 = \sup_{t, \tau \in [0, T]} \|C(t)\Phi(t, \tau)B(\tau)\Psi(\tau)\|$.

The $e^{-\lambda t}$ is multiplied at both ends of Equation (10), and $\lambda > 0$, then

$$\|e_{k+1}\|_\lambda \leq \tilde{\rho} \|e_k\|_\lambda, \quad (11)$$

where $\tilde{\rho} = \bar{\rho} + b_1 \frac{1 - e^{-\lambda T}}{\lambda} + b_2 \left(\frac{1 - e^{-\lambda T}}{\lambda} \right)^2$.

It can be inferred from condition (1) that there is $\tilde{\rho} < 1$ if λ is biggest.

Therefore $\lim_{k \rightarrow \infty} \|e_k\|_\lambda = 0$.

The proof is ended.

This paper makes the output signal sequence $u(t)$ converge uniformly optimal signal $u^*(t)$ based on PID-type iterative learning control theory. In addition, the output $y_k(t)$ conforms with the expectation trajectory

tracking of $y_D(t)$ in $t \in [0, T]$. The control process of PID-type iterative learning control algorithm is shown in Figure 2. The idea of ILC is to gradually revise imperfect control input using the error between system output and the desired trajectory and realize perfect tracking in a finite time interval.

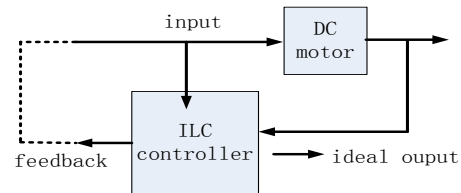


FIGURE 2 Iterative learning control

4 Simulation

To confirm the algorithm's effectiveness, the parameters of DC motor are shown in Table 1. According to the selected parameter in Formula 4, the motor's transfer function can be obtained as Equation 12.

$$G(s) = \frac{100}{s^2 + 10s + 1} \quad (12)$$

Discrete the transfer function and sampling time is 1ms. The linear system of controlled object is obtained as Equation (13).

$$\begin{cases} X(k+1) = AX(k) + Bu(k) \\ y_v(k) = CX(k) \end{cases}, \quad (13)$$

where $A = \begin{bmatrix} 1 & 0.001 \\ 0 & 0.9753 \end{bmatrix}$, $B = \begin{bmatrix} 0.0001 \\ 0.1314 \end{bmatrix}$ and $C = [1 \ 0]$.

The specific simulation parameter is $\Gamma = \begin{bmatrix} 0 & 1 \\ 0.5 & 0.6 \end{bmatrix}$, $\Phi = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and $\Psi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$.

The ideal trajectory is the $y_d(t) = \sin(3t)$. The simulation results are velocity tracking and error respectively in different iterations, and the results are as shown in Figure 3-10. The iterations are 5, 10, 20 and 30. From the results, the velocity tracking differs from the ideal track in the first iteration cycle. With the increase of iterations, the velocity tracking trajectory approaches to the ideal trajectory more and more. Meanwhile the velocity tracking error is becoming smaller and smaller.

Especially after the iterations reach twenty times, the velocity tracking trajectory nearly coincides with the ideal track. Based on the above, it can be confirmed that PID-type iterative learning control algorithm is very suitable for the velocity trailing of DC motor with repeating motion property. In addition, the more the times

of repetition is, the more the velocity trailing accuracy is, and the smaller the velocity trailing error is. It can be deduced that the motor can fully track the given expectation trajectory in certain value of iterations. Meanwhile, the control performance is verified of PID-type iterative learning control algorithm is verified by the experiments.

TABLE 1 Relevant parameters of DC motor

parameter	numerical value
K_t	0.001
K_e	0.01
R_m	0.1 Ω
L_m	10mH
J	0.001kg.m ²
B	0

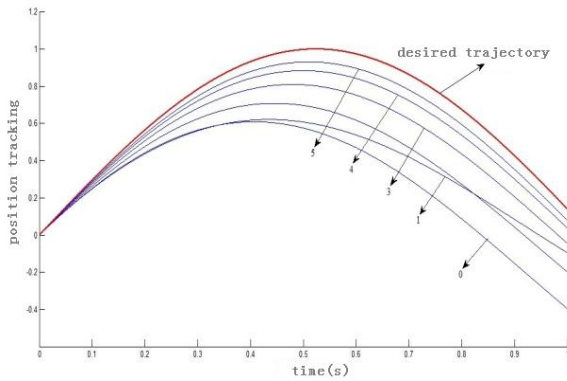


FIGURE 3 Velocity tracking at the iterations=5

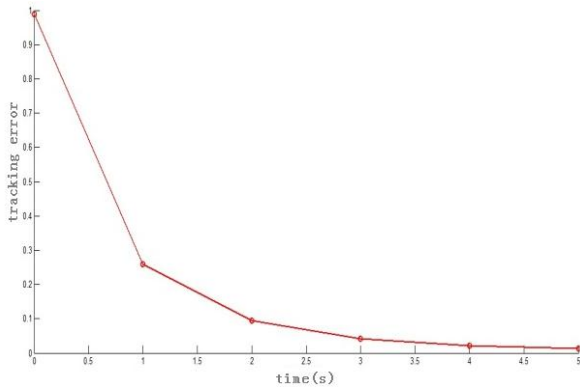


FIGURE 4 Tracking error at the iterations=5

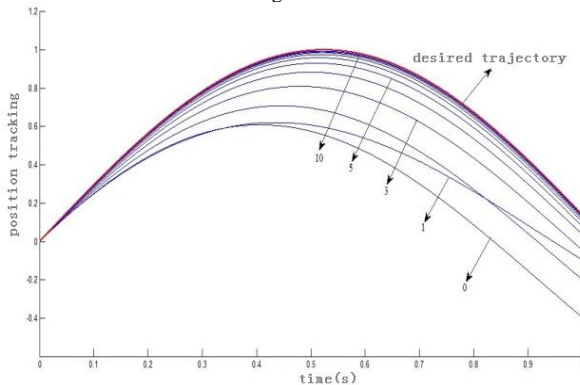


FIGURE 5 Velocity tracking at the iterations=10

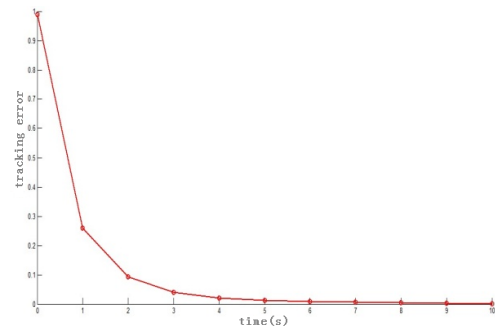


FIGURE 6 Tracking error at the iterations=10

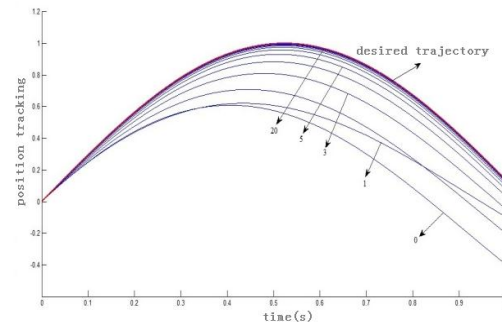


FIGURE 7 Velocity tracking at the iterations=20

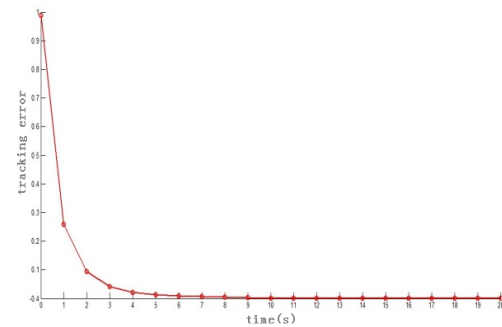


FIGURE 8 Tracking error at the iterations=20

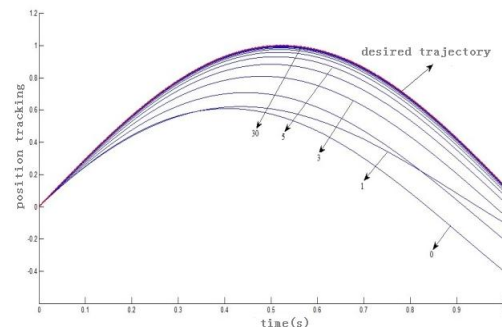


FIGURE 9 Velocity tracking at the iterations=30

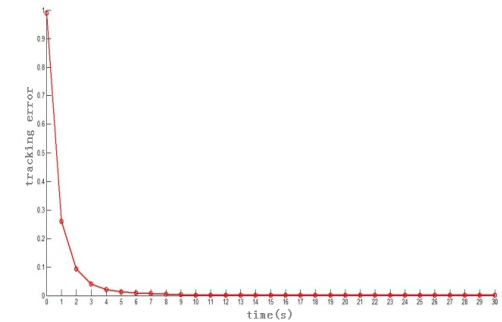


FIGURE 10 Tracking error at the iterations=30

5 Conclusion

The paper applied PID-type iterative learning control algorithm to control the velocity of DC motor with the feature of repeating motion. By the simulation results we can obtain the following conclusions:

(1) With the increase in the number of repetitive motion of the DC motor, the velocity tracking errors by a PID-type iterative learning control algorithm is getting smaller and smaller, according to the simulation shows that when the iterative learning is more than 20 times, the actual velocity and the ideal one substantially coincident.

(2) The motor can fully track the given expectation trajectory in certain value of iterations. At the same time,

it is proved that iterative learning control algorithm can ceaselessly evaluate the control performance of the system by taking full advantage of the periodicity or repeatability of the dynamic behaviour. It can amend the control signal of the system, and improve the control performance of the system. It is also revealed from simulation results that the proposed control strategy is valid and effective for the DC motor.

Acknowledgments

This work is funded by the Scientific Research Projects of Xianyang Normal University(No.11XSYK 329, No.13XSYK053, No.13XSYK055, No.13XSYK05 8).

References

- [1] Hou Li, Fan Qingwen, Huang Chengxiang 2003 *System Design of Mechanical and Electronical Integration* Beijing Higher Education Press: Beijing
- [2] Arimoto S, Kawamura S, Miyazaki F 1984 Bettering operation robots by learning *Journal of Robotic System* **1**(2) 123-40
- [3] Arimoto S, Kawamura S, Miyazaki F 1984 Bettering operation of dynamic systems by learning: a new control theory for servomechanism or mechatronic system *Proceedings of 23rd conference on decision and control* Las Vegas Nevada 1064-69
- [4] Xu J X, Tan Y 2002 Robust optimal design and convergence properties analysis of iterative learning control approaches *Automatica* **38**(11) 1867-80
- [5] Lee J H, Lee K S, Kim W C 2000 Model-based iterative learning control with a quadratic criterion for time-varying linear systems *Automatica* **36**(5) 641-57
- [6] Hatonen J, Owens D H, Feng K 2006 Basis functions and parameter optimization in high-order iterative learning control *Automatica* **42**(2) 287-94
- [7] Bien Z, Xu J X 1998 *Iterative Learning Control: Analysis, Design, Integration and Applications* Kluwer Academic Publishers: Boston USA
- [8] Yang Hong 2011 Temperature Control of Resistance Furnace Based on PID-Type Iterative Learning Control Algorithm *Journal of Shanghai University of Engineering Science* **25**(4) 351-4 (In Chinese)
- [9] Xu J X, Tan Y 2003 *Linear and Nonlinear Iterative Learning Control* Springer-Verlag: Berlin
- [10] Lee J H, Lee K S 2007 Iterative learning control applied to batch processes: an overview *Control Engineering Practice* **15**(10) 1306-18
- [11] Al-Towaim T, Barton A D, Lewin P L, Rogers E, Owens D H 2004 Iterative learning control - 2D control systems from theory to application *International Journal of Control* **77**(9) 877-93
- [12] Janssens P, Pipeleers G, Swevers J 2013 A data-driven constrained norm-optimal iterative learning control framework for LTI systems *IEEE Trans. on Control Systems Technology* **21**(2) 546-51
- [13] Saab S S 2003 Stochastic P-type/D-type iterative learning control algorithms *International Journal of Control* **76**(2) 139-48
- [14] Ratcliffe J D, Hatonen J J, Lewin P L, Rogers E, Harte T J, Owens D H 2005 P-type iterative learning control for systems that contain resonance *International Journal of Adaptive Control and Signal Processing* **19**(10) 769-96
- [15] Moore K L 2001 An observation about monotonic convergence in discrete-time, P-type iterative learning control *Proceeding of the IEEE International Symposium on Intelligent Control MX USA* 45-9
- [16] Chen C K, Hwang J 2006 PD-type iterative learning control for trajectory tracking of a pneumatic X-Y table with disturbances *JSME International Journal, Series C* **49**(2) 520-6
- [17] Chen Y Q, Moore K L 2002 PI-type iterative learning control revisited *Proceeding of the American Control Conference, Anchorage, Alaska, May* 1262-7
- [18] Park K-H, Bien Z, Hwang D-H 1999 A study on the robustness of a PID-type iterative learning controller against initial state error *International Journal of Systems Science* **30**(1) 49-59
- [19] Chien Chiang-Ju, Tai Chia-Liang 2004 A DSP Based Sampled-Data Iterative Learning Control System for Brushless DC Motors *Proceedings of the 2004 IEEE International Conference on Control Applications* 995-1000
- [20] Chien C J 2006 An adaptive PID-type iterative learning controller for nonlinear systems with non-repeatable control tasks *Transactions of the Chinese Institute of Engineers, Series A* **29**(2) 279-87
- [21] Huang Y C, Chan M, Hsin Y P, Ko C C 2003 Use of PID and iterative learning controls on improving intra-oral hydraulic loading system of dental implants *JSME International Journal, Series C* **46**(4) 1449-55
- [22] Ma F, Li C 2011 Open-closed-loop PID-type iterative learning control for linear systems with initial state error *Journal of Vibration and Control* **17**(12) 1791-7
- [23] Ali Madady 2013 An Extended PID Type Iterative Learning Control *International Journal Automation and Systems* **11**(3) 470-81
- [24] Madady Ali 2008 PID type iterative learning control with optimal gains *International Journal of Control, Automation, and systems* **6**(2) 194-203
- [25] Toscano R 2007 Robust synthesis of a PID controller by uncertain multimodal approach *Information Sciences* **177**(6) 1441-51
- [26] Liu Haishan, Chen Yuchen 2009 Simulation and Experiment Study of BLDCM PID Control System *Journal of System Simulation* **21**(16) 5157-60
- [27] Jia Hongping, Wei Haifeng 2010 Variable Structure Sliding Mode Current Control of BLDCM Micromotors **34**(2) 58-61
- [28] Zhao Ruiping 2006 *Design and Simulation of Control System for Brushless DC Motor* Master Thesis of Harbin Engineering University Harbin China

Authors



Hong Li, born in July, 1976, Xianyang, Shaanxi, P.R. China

Current position, grades: School of Information Engineering, Xianyang Normal University, China.

University studies: BS degree in computer applications technology in China. MS degree in computer applications technology from Northwest University for Nationalities in China.

Scientific interest: different aspects of control engineering, image processing and artificial intelligence.

Publications: more than 25 papers published in various journals.

Experience: 15 years, seven scientific research projects.

Based on pressure gradient model to determine leakage point in heating pipe network

Xianliang Yang*, Lianlian Jia, Songling Wang, Jiangjiang Wang

Department of Energy and Power Engineering, North China Electric Power University, Baoding City, Hebei Province, China, 071003

Received 10 December 2013, www.tsi.lv

Abstract

This paper describes a calculation method based on pressure gradient model to determine leakage point in central heating pipe network which decreases the effect of the ratio frictional resistance. In the calculation method, a pipe resistance characteristic coefficient is introduced. This characteristic coefficient is corrected by hybrid adaptive genetic algorithm. With the characteristic coefficient and the pressure value of each node in heating pipe network, the leakage point orientation in central heating pipe network can be done with the positioning analysis on the heating pipe network leak. The pressure value is calculated using the theory of graph algorithms. In this analysis, the pressure gradient is the most important analytical method. According to the data monitored by supervisory control and data acquisition (SCADA) system, the leak position can be located. The experimented results show that the method to locate the leakage point in the thermal system pipe network meets the theoretical need and the experimental one.

Keywords: Heating pipe network, Leakage location, Drag coefficient, Pressure gradient

1 Introduction

Nowadays, with the rapid development of modern heating industry in big cities, a variety of heating pipe networks have covered the subsurface of the city. But there are some problems of the heating pipe network in the heating system such as pipeline leakage. When the leakage occurred because of corrosion, natural disaster, the destructive activities of human beings and other reasons, it would lead to economic and thermal resource loss. And even more, the residents' lives would be impacted seriously [1].

At present, heating pipe network is so large that it is impossible for people to determine the pipe leakage only by person. The computer monitoring must be introduced to the heat pipe network analysis. There are many detection and location methods [2-4]. However, many of these methods have constraints. For example, the negative pressure wave method can detect the leakage of sudden. It is not accurate for the leakage point orientation in pipe network with a slow leak or a small leak. Moreover, it is much dependent on the measured value of node. Currently the model of pipeline leak detection and location is divided into transient model and homeostasis model. The homeostasis model is the leakage location by calculating the turning point of pressure gradient in leaking pipeline [5-8].

In this paper, the homeostasis model is used to locate the leakage of pipe segment. The drag coefficient (M) is used to reduce the computational difficulty of homeostasis model. Then the hybrid adaptive genetic algorithms and the graph theory calculation method are introduced to calculate the data of nodes. After the

calculation, the hydraulic model is used to derive the leak location of pipe network when the leakage is less than the amount of maximum fill water.

2 Methodology

2.1 HYDRAULIC MODEL

The pressure gradient hydraulic model shows that the pressure diagram is straight line along with the pipeline when leakage does not occur in the pipe network. However, when the pipe segment appears a leakage, the water flow before the leak point in pipe segment will increase. At the same time, in the segment, the pressure gradient is bigger and the friction is riser. On the contrary, water flow after the leak point in pipe segment will decrease. Then, the pressure gradient becomes lesser and the friction becomes lower in the later segment. Therefore, the pressure diagram in the pipe section changes from the straight line into a fold line when the leakage occurs in pipe segment. The turning point in this fold line can be determined as the leakage point of the pipe section [9]. The change of pressure diagram is shown in Figure 1.

According to this change of pressure data of each point in the pipe section, the distance from the former node of the leak segment to the leakage point can be calculated as Equation 1.

$$X = \frac{P_1 - P_2 - LR_2}{R_1 - R_2}, \quad (1)$$

* *Corresponding author* e-mail: yx115820@163.com

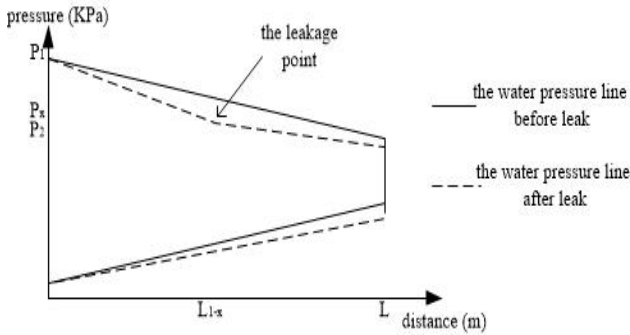


FIGURE 1 The Change of Pressure Diagram

where, X is the distance from the former node to the leakage point. P1 is the pressure of the former node of this pipe segment. P2 is the pressure of the latter node of this pipe segment. R1 is the proportion friction resistant from the former node to the leakage point. R2 is the proportion friction resistant from the leakage point to the latter node. L is the length of this pipe segment.

In order to get the distance from the former node to the leakage point, the proportion friction resistant is necessary. However, it is not always the same. In actual operation of the pipe network, the pipeline corrosion and deposition of microorganisms lead to the changes of hydraulic condition along with the pipeline. They cause the changes of the proportion friction resistant in the pipe segments. Therefore, it is necessary to consider the pressure drop of fluid in different pipeline along with the changes of proportion friction resistant [10-12].

The pressure drop of the pipe section is calculated to Equation 2.

$$\Delta P = R(L + L_d) = 0.0625 \frac{\lambda G_t^2}{\rho d^5} (L + L_d), \tag{2}$$

where: ΔP is the pressure drop of the pipe section; R is the ratio frictional resistance; λ is the coefficient of resistance along the way; G_t is the flow rate of fluid in the pipe segment, t/h; d is the inner diameter of the pipe, m; L is the length of the pipe section, m; L_d is the local resistance equivalent length in the pipeline, m. R is expressed as Equation 3.

$$R = 0.0625 \frac{\lambda G_t^2}{\rho d^5}. \tag{3}$$

The rule that R is proportional to λ is obvious in Equation 3. During the actual operation of heat pipe network, the impact of hot water's quality in the pipe network, the useful life of heat pipe and the coefficient of resistance along the pipeline will change with the pipeline conditions. In order to determine the leakage point, a drag coefficient M of heat pipe network is introduced as Equation 4:

$$M = \frac{0.0625\lambda}{\rho d^5}. \tag{4}$$

Then, the Equation 3 could be expressed as Equation 5:

$$R = MG_t^2. \tag{5}$$

The introduction of drag coefficient M makes the calculation of ratio frictional resistance more simple and accurate.

2.2 DRAG COEFFICIENT

The drag coefficient M is used to analyse the leakage point of pipe network. During this analysis, a straight pipe section with pipe branch is selected to be a case study. It has the same diameter, material, and applicable life. The nodes' positions on this pipe section are shown in Fig.2. In this pipe section, the pressure of node 1 is P1 and the flow of it is Q. The pressure of node 2 is P2 and the flow in the branched pipe is q. The pressure of node 3 is P3 and the flow is Q-q. The length of tube segment 1~2 is L1. And the length of tube segment 2~3 is L2. The friction of the tube segment 1~2 is R1, and the friction of the tube segment 2~3 is R2.

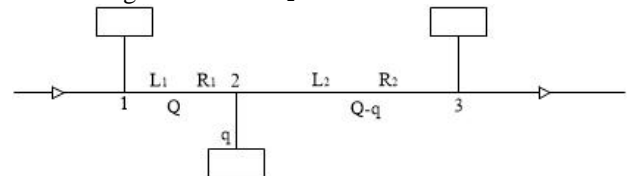


FIGURE 2 the Pipe Diagram

According to Equation 2 to Equation 5, the following equation is obtained:

$$R = \frac{\Delta P}{LP + L_d} = MG_t^2. \tag{6}$$

Then the drag coefficient M could be calculated.

$$M = \left(\frac{\sqrt{R_1} - \sqrt{R_2}}{q} \right)^2 = \left(\frac{\sqrt{\frac{P_1 - P_2}{L_1 + L_{d1}}} - \sqrt{\frac{P_2 - P_3}{L_2 + L_{d2}}}}{q} \right)^2. \tag{7}$$

In the Equation 7, when the pressure data and flow data of node 1, 2, 3 are known, the drag coefficient M of this pipe segment could be calculated.

2.3 HYBRID ADAPTIVE GENETIC ALGORITHM MODEL

In actual heat pipe network running, the drag coefficient M is not always constant. It must be corrected along with the changes of working conditions of heat pipe network.

Herein the hybrid adaptive genetic algorithm is introduced to correct M.

A characteristic of genetic algorithm is that the next search information could be got with the objective function only. The use of objective function is based on the evaluation of the fitness of individual. The individuals which have a high fitness have a high probability to inherit to the next generation. The individuals which have a low fitness have a lower probability to inherit to the next generation. The measure individual fitness is called to fitness function, and the fitness function is objective function in optimization problem [13-15]. In this analysis, the following fitness function is defined to:

$$\min F(S) = \sum_{i=1}^N (P_{ic} - P_{is})^2. \tag{8}$$

The constraint conditions in objective function include the following formulas:

$$\begin{cases} \sum G_{m,ij} + Q_i = 0 & (i=1,2,\dots,N) \\ P_{ij} = S_{ij} |G_{m,ij}| G_{m,ij} - dH_{ij} & (i=1,2,\dots,N), \\ S_{\min} \leq S_{ij} \leq S_{\max} & (i=1,2,\dots,N) \end{cases} \tag{9}$$

wherein:

- P_{ic} is the pressure value of detection point measured in the pipe network system, Pa;
- P_{is} is the pressure value of monitoring point calculated by the graph theory, Pa;
- $G_{m,ij}$ is the flow in the pipes connected with the node i, m^3/h ;
- Q_i is the leak flow of node i, m^3/h ;
- N is the number of node;
- dH_{ij} is the pump head connected to the node i, m;
- S_{\min} is the lowest limit of the coefficient of drag characteristics of each pipe section in the pipe network vector, $Pa \cdot h^2/t^2$;
- S_{\max} is the supreme limit of the coefficient of drag characteristics of each pipe section in the pipe network vector, $Pa \cdot h^2/t^2$.

In these constraints, the coefficient of drag characteristics of each pipe section in the heat pipe network can be corrected within engineering error when the balance of heat pipe network hydraulic can be kept. The deviation between the node pressure data, which is measured and calculated by figure in the pipe network can be decreased.

2.4 POSITIONING OF LEAKAGE POINT

According to the M, which is corrected by hybrid adaptive genetic algorithm, a homeostasis model is established for the pipeline leakage detection. The model

is used to find out the leakage point in the pipe network, which is a single pipe section. There is a leakage point between the node 2 and the node 3 in this pipe section, which is shown in Figure 3.

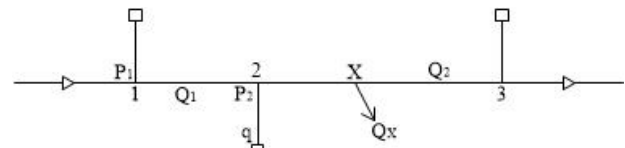


FIGURE 3 the leak of tube section 2 to 3

In Figure 3, the flow rate of node 3 is Q_2 . The leak point is X. The amount of leakage is Q_x . The flow rate of node 1 can be calculated as $Q_1 = Q_x + Q_2 + q$. The position of the X can be determined as long as the L_{2-X} or L_{X-3} is calculated. The calculated process is as follows:

$$L_{2-X} = \frac{P_2 - P_3 - MQ_2^2 L_{2-3}}{M \left[(Q_1 - q)^2 - Q_2^2 \right]}, \tag{10}$$

where, L_{2-X} represents the length that includes all the resistance loss between the node 2 to the leakage point. L_{X-3} represents the length that includes all the resistance loss between the leakage points to the node 3. L_{2-3} represents the length that includes all the resistance loss between node 2 to node 3 when the pipe do not occur leak.

3 Case studies

In order to verify the accuracy of the pressure gradient method, a heating pipe network leak detection system is used. This system is used to simulate the running conditions of actual heating pipe network. The photograph of it is showed in Figure 5. Its running sketch is shown as Figure 4. The SCADA in this system could monitor the data of the constant pressure point, and the pressure of import and export of circulating water pump. At the same time the total flow data in heat export and the flow data of individual (or heat transfer station) and the pressure data of heat exchanger could be monitored by SCADA.

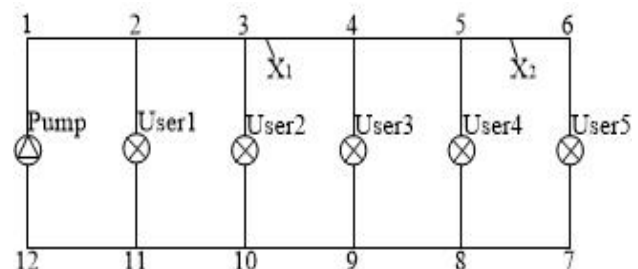


FIGURE 4 The running sketch



FIGURE 5 The photograph of pipe network

The data, which is used in calculation process is gotten after the experiment. The data of constant pressure point in the experimental pipe network system is 120595Pa. The first step researchers let the leakage is on point X₁ and the flow of this point is 0.491m³/h. And then researchers record the data of monitoring points. The second step researchers let the leakage point is X₂ and the

flow of it is 0.489m³/h. The data to be recorded are shown in Table1. With the experimental data, the hybrid adaptive genetic algorithm is used to correct the drag coefficient M of the pipe network. After that, the graph theory is used to calculate the pressure data of nodes. The results are shown in Table 1, and the computation of leakage point is shown in Table 2.

TABLE 1 The calculate data of node pressure

Node number	Actual measurement value (Pa)	Graph theory that has not been corrected calculated (Pa)	Error rate	Calculated value of the graph theory after the genetic algorithm Corrected (Pa)	Error rate
3	342301	337270	1.47%	338540	1.10%
4	331651	327530	1.24%	327870	1.14%
5	311185	307201.8	1.28%	308695.5	0.8%

TABLE 2 The calculate value of the leakage point

The characteristics breakdown of the tube network M		L _{3-X1} (m)	Error rate
operating conditions not optimized	1.397493	23.3557	42.68%
operating conditions optimized	1.5794	9.289978	6.707%
The characteristics breakdown of the tube network M		L _{4-X2} (m)	Error rate
operating conditions not optimized	1.4967	31.5594	32.91%
operating conditions optimized	1.6882	43.2569	8.76%

In the Table 2 the error rate of operating condition without optimizing is big; on the other hand it with optimizing is low. From the Table 1 and Table 2, it can be seen that the calculation accuracy of the node's pressure is improved after using the corrected drag coefficient M of pipe network. Thus, the error that caused by the pressure gradient model which used in leak point positioning is decreased. Therefore, the method in this paper is effective in the leak location of heating pipe network. Moreover, it shortens the time to check the leak points and reduces the heat wasting.

4 Conclusions

This paper introduces a pipe drag coefficient M. In the calculation process, the drag coefficient M, which is corrected by hybrid adaptive genetic algorithm, avoids the influencing factors of the coefficient of drag

characteristic, the pipe frictional, resistance coefficient and the inner diameter in this pipe. It makes the calculation of leak location easier and more accurate. The model of pressure gradient is proved to be effective in leakage location in heating pipe with the heating pipe network leakage detection system. Compared with the fact, the average error rate of the calculation in experiment is 7.7335%. This error meets project need.

Acknowledgement

This work was financially supported by the Fundamental Research Funds for the Central Universities (13MS99) and Beijing Natural Science Foundation (3122028), Special Funds for Excellent Doctoral Dissertation of Beijing (20121001901) and Hebei Natural Science Foundation (E2012502002).

References

- [1] Bruegger P, Pallotta V, Hirsbrunner B 2009 Optimizing heating system management using an activity-based pervasive application *Journal of Digital Information Management* 6(7) 327-35
- [2] Lin Weiguo, Zheng Zhishou 2006 Research on Pipeline Leak Detection Based on Dynamic Pressure Signal *Chinese Journal of Scientific Instrument* 27(8) 907-10
- [3] Andrew F C, Pedro L Bryan W K 2009 A Selective Literature Review of Transient-based Leak Detection Methods *Journal of Hydro-Environment Research* 4(2) 212-27
- [4] Ge C H, Yang H Y, Ye H, et al 2009 A Fast Leak Locating Method Based on Wavelet Transform *Tsinghua Science & Technology* 14(5) 551-5
- [5] Yang J, Wen Y M, Li P 2008 Leak Location Using Blind System Identification in Water Distribution Pipelines *Journal of Sound and Vibration* 3(10) 134-148
- [6] Sun Liang, Wang Jianlin 2012 Release Transient Modeling Based Leak Detection and Location Method for Pipelines *Journal of Basic Science and Engineering* 20(1) 159-68
- [7] Zhang X J 1993 Statistical Leak Detection in Gas and Liquid Pipelines *Pipes & Pipelines International* 8(2) 26-9
- [8] Sun Liang, Wang Jianlin 2010 Research on the Method for Leak Location in Gas Pipeline Based on Ateady-state Model *Chinese Journal of Scientific Instrument* 31(3) 565-70
- [9] Wang Guizeng, Ye Hao 2012 *Leak Detection and Location of Fluid Transporting pipeline* Tsinghua University Press: Beijing (in Chinese)
- [10] Li Deying, Xu Wenfa 2006 *Heating Project China*, the Building Industry Press of China: Beijing (in Chinese)
- [11] Long Tianyu, Cai Zengji 2004 *Fluid Mechanics China*, the Building Industry Press of China: Beijing. (in Chinese)
- [12] Fu Xiangzhao, Mao Qingxi 2005 *Fluid Transmission and Distribution Network China*, the Building Industry Press of China: Beijing (in Chinese)
- [13] Li Minqiang 2002 *the Basic Theory and Applications of Genetic Algorithms China* Science and Technology Press of Beijing: Beijing (in Chinese)
- [14] Yang Xianliang, Li Tong, Wang Songling, Wu Zhengren 2012 The research on the resistance characteristic coefficient in heating system based on the improved genetic algorithm *Applied Mechanics and Materials* 170(173) 2587-91
- [15] Palanivel R, Mathews P K, Murugan N 2011 Development of mathematical model to predict the mechanical properties of friction stir *Journal of Engineering Science and Technology Review* 1(4) 25-31

Authors	
	<p>Yang Xianliang, born in May, 1975, Baoding County, Hebei Province, P.R. China</p> <p>Current position, grades: The experiment centre director of School, North China Electric Power University, China.</p> <p>University studies: B.Sc. in department of urban construction engineering from Hebei Institute of Architecture and Civil Engineering College in China. M.Sc. from North China Electric Power University in China.</p> <p>Scientific interest: heating energy saving.</p> <p>Publications: more than 20 papers published in various journals.</p> <p>Experience: 13 years, 5 scientific research projects.</p>
	<p>Jia Lianlian, born in February, 1989, Baoding County, Hebei Province, P.R. China</p> <p>University studies: B.Sc. in department of civil engineering from Tangshan College in China.</p> <p>Scientific interest: heating energy saving.</p> <p>Publications: 2 papers published in various journals.</p>
	<p>Wang Songling, born in November, 1954, Baoding County, Hebei Province, P.R. China</p> <p>Current position, grades: Professor of School, North China Electric Power University, China.</p> <p>University studies: B.Sc. from North China Electric Power University in China.</p> <p>Scientific interest: thermal equipment and safety and economic operation of large rotary machinery and Fluid dynamics theory and application.</p> <p>Publications: more than 70 papers published in various journals.</p> <p>Experience: 30 years, more than ten scientific research projects.</p>
	<p>Wang Jiangjiang, born in January, 1979, Shijiazhuang County, Hebei Province, P.R. China</p> <p>Current position, grades: Associate Professor of School, North China Electric Power University, China.</p> <p>University studies: B.Sc. in department of automation from North China Electric Power University in China. M.Sc. from North China Electric Power University in China.</p> <p>Scientific interest: heating energy saving.</p> <p>Publications: more than 30 papers published in various journals.</p> <p>Experience: 8 years, 2 scientific research projects.</p>

External locating of moving targets for 3D IMRT using parallax method

Jian-Ning Han*, Peng Yang, Lu Zhang

School of Information and Communication Engineering, North University of China, 3 Xue Yuan Road, Taiyuan, China

Received 26 March 2014, www.tsi.lv

Abstract

Due to the complexity of intensity modulated radiotherapy, it is difficult to meet the current situation of treatments which require real time locating of moving target. In this paper, we propose a new method based on parallax method using external label to locate the moving target. A column test-piece is used to simulate the human body structure for data analysis. A defect model for human body target is implanted inside the test-piece, which is the moving target. Based on the parallax method, the depth of the defect can be obtained by using the two images captured before and after the test-piece moves a short distance. The possible errors which affect the test results are analysed. The effects of errors can be reduced by adjusting the system parameters. The results show that the parallax method is a simple but efficient approach, which can be used for locating moving target in intensity modulated radiotherapy.

Keywords: Intensity modulated radiotherapy, Moving target location, Parallax method, Three-dimensional location, External Label

1 Introduction

Nowadays, intensity modulated radiotherapy has been able to realize the appropriate radiotherapy for a static target according to its height [1]. However, while the target moves during the process of radiotherapy, the blurring effects and the interaction between fringes of image will occur [2]. Therefore, it is difficult to arrange an appropriate dose arrangement for the moving target due to the dynamic state of human body structure, for instance, the tumor target in the human organs such as chest, stomach or prostate of a patient.

Recently, there are some studies on the target real time locating method. For instance, accurate real time dynamic information of the target can be obtained by implanting the signal responder unit inside the target. However, it also increases the complexity. More studies focus on the metal label's image inside the target to detect real time location. Until now, on the latest conventional accelerator, only 1KV ray source and one image device are installed vertically on the treatment ray bundle. Although the stereoscopic image location of the target can be achieved by using MV ray and KV ray treatment in some situation, the shadow of the label on EPID within the ray scope of the intensity modulated radiotherapy will often be sheltered from the MLC, which will have an effect on the location. However, the contrast ratio of MV image is far less than KV image, which makes it difficult to identify inner label in practice [3].

More recently, additional dimensional imaging equipment are used to process real time moving target location in some medical centres. For example, IRIS system (integrated radiotherapy imaging system) [8, 9]

which can be installed on the accelerator's rack and rotates with the rack or RTRT system (Real-time tumor-tracking system). But these equipment not only expensive but also bring overdose imaging for patients during treatment. Considering these reasons, lots of studies try to use one direction LV X-ray imaging to achieve the real-time location for moving targets. Because the method is simple to apply and no need to use extra equipment, the dose for imaging of can be reduced. In this paper, we present a new three-dimensional locating approach for moving targets by using external label, which combines the features of moving targets with the theory of parallax method. A column test-piece with inner defects is used to simulate the human body targets. This real time locating method for the moving target can be used to get information for gating radiotherapy or tracking radiotherapy [4, 5].

2 System model design for radiation examination

2.1 IMAGING SYSTEM

Our method uses a column test-piece to represent the organization of human body. The internal defects of cylinder simulate the target area of the human body. The cylinder movement represents the human organ motion, The single direction X-ray system is shown Figure 1. The medical X-ray system is shown in Figure 2, Figure 3 shows the defects in cylinder which are very similar to human target area [6].

* *Corresponding author* e-mail: hanjn46@nuc.edu.cn

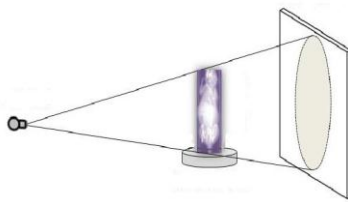


FIGURE 1 Medical Single direction X-ray system

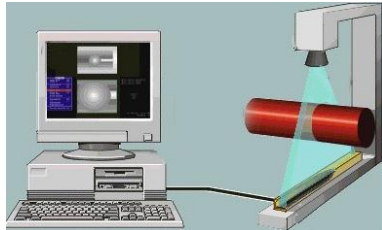


FIGURE 2 The Simulation of medical X-ray system

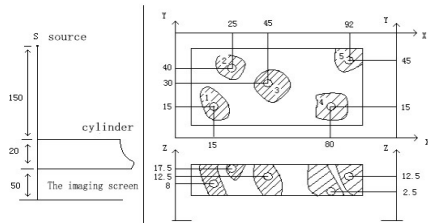


FIGURE 3 The cylinder with internal defects

2.2 MATHEMATICAL MODEL

We use the parallax method for defect locating. It is based on the straight propagation characteristics of radiation and the theory of triangle line similarity. A point (an object) at different locations of the two planes has been used to determine the spatial position. The schematic diagram is shown in Figure 4 [7].

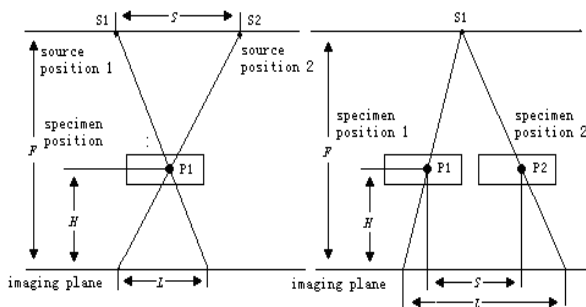


FIGURE 4 Parallax method principle diagrams

The left part of Figure 4 shows the locating method that the ray source moves. Ray source moves from S1 position to S2 position, the corresponding image point P1 has a displacement of L [8].

Using the triangle similarity relation, we can get:

$$H = LF / (S + L), \tag{1}$$

wherein:

- H*-the distance from defects to imaging plane;
- F*-the distance from ray source to imaging plane;
- S*- displacements between ray source exposure;

L-parallax, that is the displacement of defect projection.

Among them, *F*, *S*, *L* can be directly measured or indirectly measured by other methods and the defect depth can be obtained the quantity into Equation (1).

The right part of Figure 2 shows the locating method that specimen moves and ray source is motionless. Using the geometry relations:

$$H = F * (1 - S / L), \tag{2}$$

wherein:

H, *F* are consistent with an earlier statement;

S-displacements between ray source exposure;

L-parallax, that is the displacement of defect projection;

The final result of two methods are the same. Either method can be used for locating moving targets.

3 Defects in three-dimensional location

The schemes of defect location can be divided into direct measurement method, single labelled point method and multiple labelled point method. These three methods are based on the principle of parallax. Among them, direct measurement method is the simplest one. It applies the parallax principle model directly and obtains the defect under one coordinate system's absolute coordinate. Because the coordinate is not direct-viewing, so the method is impractical. Single labelled point method is quite easy and practical, it based on a marked point which is put on the specimen surface as a reference point. Finally we obtain the defect point coordinates relative to the reference point. It is intuitive and accurate, and it is the method we will adopt. The principle of double labelled point method and the single labelled point method are same, but they are complicated. Here we will mainly introduce the single labelled point method [9].

As is shown in Figure 4, putting a reference point on the surface of an object, we can use the equation to get the three-dimensional coordinates of M:

$$\begin{cases} X_M = L_M x_M / S \\ Y_M = L_M y_M / S \\ Z_M = F * (1 - S / L_M) \end{cases}, \tag{3}$$

$$\begin{cases} X = L_D x_D / S - L_M x_M / S \\ Y = L_D y_D / S - L_M y_M / S \\ Z = F * (1 - S / L_D) - F * (1 - S / L_M) \end{cases}. \tag{4}$$

After simplification, we have:

$$\begin{cases} X = (L_D x_D - L_M x_M) / S \\ Y = (L_D y_D - L_M y_M) / S \\ Z = FS * (1 / L_M - 1 / L_D) \end{cases}. \tag{5}$$

From the above equations, we know that we need to determine the origin of coordinate in the process of measuring the x_D, y_D, x_M, y_M . But the process of subtraction can greatly eliminate common errors, and the coordinates got by this method are intuitive and easy to

understand. We can use the laser illuminates along the rays to imaging screen and the screen image point of the laser can be used as the origin of the coordinate system [10].

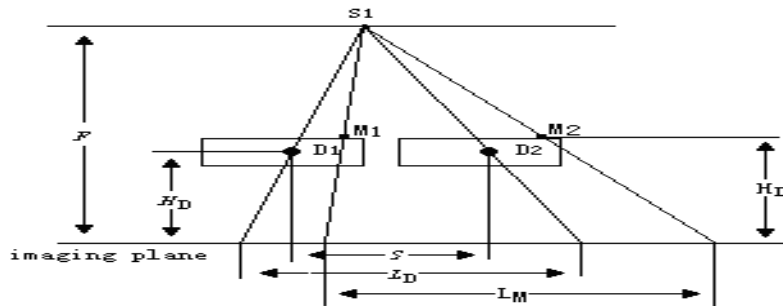


FIGURE 5. Single point external labelled defect location plans

4 Error analysis for defect location

4.1 THE UNCERTAINTY ESTIMATION OF INDIRECT MEASUREMENT

Because the three-dimensional coordinates of defect (x, y, z) are indirect measurement result, which cannot be measured directly, we firstly introduce the uncertainty estimation of indirect measurement.

If indirect measurement value N is the function of independent direct values of $x, y, z: N = f(x, y, z)$. And x, y, z contain errors, so N must contain errors. When the error is small, according to the Taylor series expansion, the uncertainty of each direct measurement result can be obtained by the uncertainty of indirect measurement result, which is called synthesis of uncertainty. There are two basic uncertainty synthesis methods [11-12].

If $\sigma_x, \sigma_y, \sigma_z$ are standard deviation of x, y, z , σ_N , the standard deviation of N is given by the following equation:

$$\sigma_N = \sqrt{\left(\frac{\partial f}{\partial x} \partial x\right)^2 + \left(\frac{\partial f}{\partial y} \partial y\right)^2 + \left(\frac{\partial f}{\partial z} \partial z\right)^2} \tag{6}$$

It turns out that no matter what probability distribution x, y, z is, this equation is tenable and this method is called the square root synthesis of standard deviation.

If e_x, e_y, e_z are error range of x, y, z , we can obtain e_N (the error range of N) through the arithmetic synthesis method:

$$e_N = \left| \frac{\partial f}{\partial x} \right| e_x + \left| \frac{\partial f}{\partial y} \right| e_y + \left| \frac{\partial f}{\partial z} \right| e_z \tag{7}$$

It is more reliable using arithmetic synthesis method when we estimate error range, although it often tends to be conservative [13].

4.2 UNCERTAINTY ANALYSIS OF THE PARALLAX METHOD

In direct measurement method, the standard deviation of X_D, Y_D, Z_D are $\sigma_{x_D}, \sigma_{y_D}, \sigma_{z_D}$ list as follows:

$$\begin{aligned} \sigma_{x_D} &= \sqrt{\left(\frac{\partial f}{\partial L_D} \partial L_D\right)^2 + \left(\frac{\partial f}{\partial x_D} \partial x_D\right)^2 + \left(\frac{\partial f}{\partial S} \partial S\right)^2} \\ \sigma_{y_D} &= \sqrt{\left(\frac{\partial f}{\partial L_D} \partial L_D\right)^2 + \left(\frac{\partial f}{\partial y_D} \partial y_D\right)^2 + \left(\frac{\partial f}{\partial S} \partial S\right)^2} \\ \sigma_{z_D} &= \sqrt{\left(\frac{\partial f}{\partial F} \partial F\right)^2 + \left(\frac{\partial f}{\partial S} \partial S\right)^2 + \left(\frac{\partial f}{\partial L_D} \partial L_D\right)^2} \end{aligned} \tag{8}$$

and can be simplified as:

$$\begin{aligned} \sigma_{x_D} &= \sqrt{\left(\frac{x_D}{S} \partial L_D\right)^2 + \left(\frac{L_D}{S} \partial x_D\right)^2 + \left(\frac{L_D x_D}{S^2} \partial S\right)^2} \\ \sigma_{y_D} &= \sqrt{\left(\frac{x_D}{S} \partial L_D\right)^2 + \left(\frac{L_D}{S} \partial y_D\right)^2 + \left(\frac{L_D x_D}{S^2} \partial S\right)^2} \\ \sigma_{z_D} &= \sqrt{\left((1 - S / L_D) \partial F\right)^2 + \left(\frac{F}{L_D} \partial S\right)^2 + \left(\frac{FS}{L_D^2} \partial L_D\right)^2} \end{aligned} \tag{9}$$

error range:

$$\begin{aligned}
 e_{x_D} &= \left| \frac{\partial f}{\partial L_D} \right| e_{L_D} + \left| \frac{\partial f}{\partial x_D} \right| e_{x_D} + \left| \frac{\partial f}{\partial S} \right| e_S \\
 e_{y_D} &= \left| \frac{\partial f}{\partial L_D} \right| e_{L_D} + \left| \frac{\partial f}{\partial y_D} \right| e_{y_D} + \left| \frac{\partial f}{\partial S} \right| e_S \\
 e_{z_D} &= \left| \frac{\partial f}{\partial F} \right| e_F + \left| \frac{\partial f}{\partial S} \right| e_S + \left| \frac{\partial f}{\partial L_D} \right| e_{L_D}
 \end{aligned}
 \tag{10}$$

and finally we have:

$$\begin{aligned}
 e_{x_D} &= \left| \frac{x_D}{S} \right| e_{L_D} + \left| \frac{L_D}{S} \right| e_{x_D} + \left| \frac{L_D x_D}{S^2} \right| e_S \\
 e_{y_D} &= \left| \frac{y_D}{S} \right| e_{L_D} + \left| \frac{L_D}{S} \right| e_{y_D} + \left| \frac{L_D y_D}{S^2} \right| e_S \\
 e_{z_D} &= \left| 1 - S / L_D \right| e_F + \left| \frac{F}{L_D} \right| e_S + \left| \frac{FS}{L_D^2} \right| e_{L_D}
 \end{aligned}
 \tag{11}$$

From the above two equations, we can see that the error of x_D is introduced by L_D, x_D, S . Its error transfer functions can be shown as $\frac{x_D}{S}, \frac{L_D}{S}, \frac{L_D x_D}{S^2}$. In the process of measuring x_D , we properly identify the distance from ray source to the object and the distance from object to the imaging screen. Make sure the two values were comparable, and there is no big value in three errors transfer coefficient. Thus, in the process of measuring x_D , each component's influence on the measurement results are within the normal range. In a similar way, the error of y_D is introduced by L_D, x_D, S . Its error transfer coefficient can be shown as $\frac{y_D}{S}, \frac{L_D}{S}, \frac{L_D y_D}{S^2}$, and there is also no big component, so the process of measuring y_D can also be normal. The error of z_D is introduced by F, S, L_D , its error transfer coefficient can be shown as, $1 - S / L_D, F / L_D, FS / L_D^2$. Due to the companionability of S and L_D , the first component is not seriously amplified. Between the two components, the value of F is

larger than the other parameters value. Therefore, the measured results are mainly influenced by S, L_D .

From the above analysis, we can see that the existence of error propagation makes measurement affected by the error of the measurement components. Here we should pay attention to z_D , namely the deriving of defect depth which is greatly affected by S and L_D . If the value of $\frac{F}{L_D}$ and $\frac{FS}{L_D^2}$ are lager, we should reduce the two rations appropriately in order to reduce the error. But under the influence of geometric sharpness of imaging systems, the values of $\frac{F}{L_D}, \frac{FS}{L_D^2}$ must be not too small. So we can only try to improve the accuracy of S and L_D . S is the moving distance of test specimen, and is influenced by mechanical platform. The precision of the platform we used is 0.005mm. If F is about 1000mm, L_D is about 30mm, then $\left| \frac{F}{L_D} \right| e_S$ is 0.16mm [14-15].

5 Analysis of experimental images and data

As shown in Figure 3, the distance from point to the workpiece is 150mm, the distance from the workpiece to the imaging screen is 50mm, the thickness of workpiece itself is 20mm, so we can know that F is 220mm. The simulation of workpiece is 100x50x20mm. There are five point defects. We will locate the point defects with the parallax method, and make a comparison with the original coordinates.

Errors are added in simulations in order to simulate the real situation. In Figure 6, the left image is for the first location image, the middle one is after moving the point source, and Figure 6(c) is the composite image by matching two images. The ray source moves along the X-axis, so there is only the parallax between the composite images along the x-axis. Therefore, the 3 dimensional coordinates of the defect can be calculated according to the parallax images and the mathematical model.

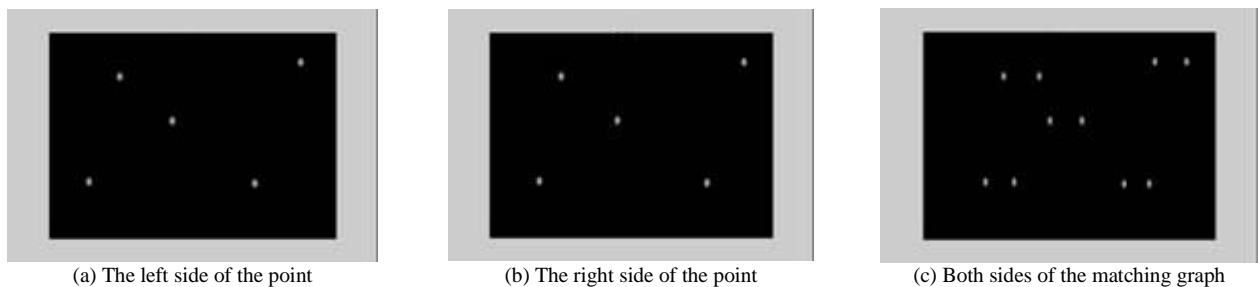


FIGURE 6 The simulation of defect projecting

Table 1 shows the comparison between the two coordinators when the error of point source movement is

0.2mm, image movement measurement error has 2 pixels, and measuring focal distance error is 0.2mm.

TABLE 1 The correlation table of seeking coordinate and original coordinate

former	X(mm)	Y(mm)	Z(mm)	calculate	X(mm)	Y(mm)	Z(mm)
1	15	15	8	1	15.102	15.102	8.00087
2	25	40	17.5	2	25.105	40.00084	17.501
3	45	30	12.5	3	45.120	30.00064	12.5009
4	80	15	2.5	4	80.152	15.00034	2.50082
5	92	45	12.5	5	92.153	45.00097	12.5009

In the experiment, the major errors are the relative parameters errors in the testing and the systematic error in application of the system. The latter is mainly the error caused by the radio source.

Figure 7 shows the relationship between H and k=S/L when ΔF is fixed as 0.2mm, 0.5mm, 0.8mm, and 1.0mm respectively. ΔH can be expressed as: ΔH=ΔF/(k+1). It can be seen that ΔH is greatly reduced as k increases.

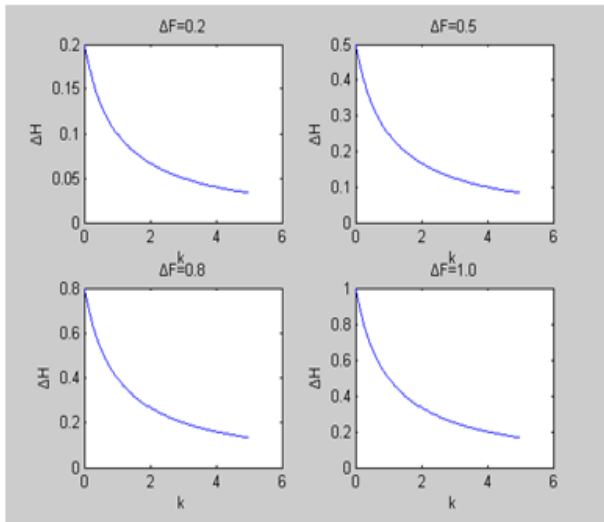


FIGURE 7 The relationship between k and ΔH when fixing ΔF

Figure 8 shows the relationship of k and ΔH when S=0.2mm, 0.5mm, 0.8mm, 1.0mm. $\Delta H = k \cdot F \cdot \Delta S / S / (k+1) / (k+1 + \Delta S / L) \approx \Delta H = k \cdot F \cdot \Delta S / S / (k+1)^2$. It can be seen that the influence of ΔS on H is affected by

the value of k. So k can be appropriately increased to reduce the influence of ΔS on the final measurement of H. Compared with the former two testing errors, the error caused by ΔL is pixel level, which can be neglected.

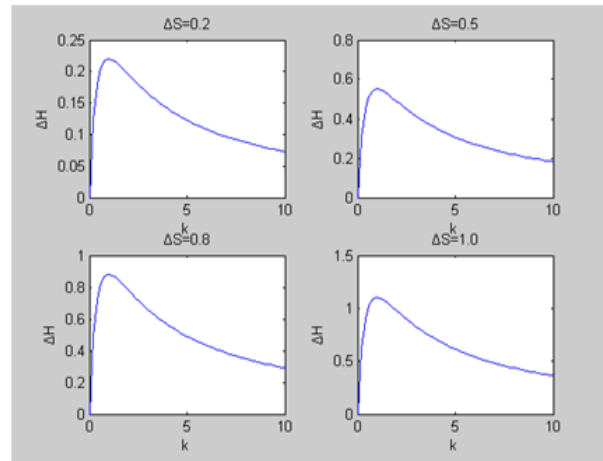


FIGURE 8 The relationship between k and ΔH when fixing ΔS

Table 2 is the coordinates when ΔF=0.5mm and ΔS=0.5mm. It can be observed that the influence of ΔF is smaller than the one of ΔS. Considering the practical testing precision can reach 0.2mm, the practical error will be smaller than that shown in the table.

TABLE 2 ΔF and ΔS on the result of measurement

insertΔF	X(mm)	Y(mm)	Z(mm)	insertΔS	X(mm)	Y(mm)	Z(mm)
1	14.904	14.904	8.132	1	15.139	15.139	7.776
2	24.898	39.869	17.654	2	25.176	40.222	17.236
3	44.858	29.905	12.642	3	45.227	30.185	12.256
4	79.762	14.955	2.620	4	80.190	15.035	2.304
5	91.709	44.858	12.642	5	92.260	45.1279	12.256

6 Conclusions

In conclusion, the parallax method is used in the external label for locating moving targets in 3D IMRT. The theoretical model based on parallax method theory is discussed and errors of parameters in measurement are analysed in detail. The images of the external labelled points at different positions are acquired in the experiment with moving ray source and fixed the workpiece. The error of defect location is about 0.2mm when the precision of testing workbench is 0.005mm,

and the distance from the ray source to image plane is 1000mm. This accuracy of measurement can meet the needs for 3D IMRT applications. Comparing with the traditional radiograph examination method, the external label method based on parallax theory is more flexible, and precise to be applied in practice.

Acknowledgements

This project was sponsored in part by The National Natural Science Foundation of Chinese NO.61302159.

References

- [1] Bertuzzi A, Bruni C, Papa F 2013 Optimal solution for a cancer radiotherapy problem *Journal of Mathematical Biology* **66**(1-2) 311-49
- [2] Van Herk M 2004 Errors and margins in radiotherapy *Seminars in Radiation Oncology* **14**(1) 52-64
- [3] Gilliom J, Monahan T 2012 *Supervision: an introduction to the Surveillance Society* University of Chicago Press: Chicago
- [4] Wiersma R D, Mao W, Xing L 2008 Combined kV and MV imaging for real-time tracking of implanted fiducial markers *Medical Physics* **35**(4) 1191-8
- [5] DeKay S H 2012 Interpersonal communication in the workplace: A largely unexplored region *Business Communication Quarterly* **75**(4) 449-50
- [6] Kieffer J C, Krol A, Jiang Z 2002 Future of laser-based X-ray sources for medical imaging *Applied Physics B* **74**(1) 75-81
- [7] Banerjee K, Dunn W L 2007 On X-ray back-scattering to detect hidden cracks in multi-layer structures *Applied Radiation and Isotopes* **65**(2) 176-7
- [8] Liu C, Lee J, Jing J 2010 Motions of hard x-ray sources during an asymmetric eruption *The Astrophysical Journal Letters* **721**(2) 193-8
- [9] Silverman D 2013 *Doing qualitative research: A practical handbook* SAGE Publications Limited: London.
- [10] Liao T W 2009 Improving the accuracy of computer-aided radiographic weld inspection by feature selection *NDT & E International* **42**(4) 229-39
- [11] FU Li-qin, HAN Yan 2008 Study on Stereo Matching Technique for Digital Radiography Images *Journal of Image and Graphics* **13**(12) 2363-7 (In Chinese)
- [12] da Silva R R, Mery D 2007 State-of-the-Art of weld seam inspection by radiographic testing: part i-image processing *E-Journal of Nondestructive Testing and Ultrasonics* **12**(4) 1-9
- [13] CHEN Ming, MA Yuezhou, CHEN Guang 2007 Weld defects detection for X-ray linear array real-time imaging *Transactions China Welding Institution* **28**(6) 81-4 (in Chinese)
- [14] Sim L M, Tan T C, Mon A A 2009 X-ray absorption-based technique to measure the thickness of multilayered structures *NDT & International* **42**(4) 291-6
- [15] Dahai R, Zheng Y, Changku, S 2000 Automatic Analysing System of X-ray Real-time Radiography for Welds *Transactions China Welding Institution* **21**(1) 60-3 (in Chinese)

Authors	
	<p>Jian-ning Han, born in December, 1980, China County, Shanxi Province, P.R. China</p> <p>Current position, grades: Lecturer of School of Information and Communication Engineering, North University of China, China. University studies: B.Sc. in Electronic information engineering from North University of China in China. M.Sc. from North University of China in China. Scientific interest: Acoustic signal processing, Graphics processing Publications: more than 7 papers published in various journals. Experience: teaching experience of 11 years, 5 scientific research projects.</p>
	<p>Peng Yang, born in July, 1988, China County, Shanxi Province, P.R. China</p> <p>Current position, grades: student of School of Information and Communication Engineering, North University of China, China. University studies: B.Sc. in Electronic information engineering from North University of China in China. M.Sc. from North University of China in China. Scientific interest: Acoustic signal processing. Experience: researching experience of 2 years, 1 scientific research project.</p>
	<p>Lu Zhang, born in July, 1988.07, China County, Shanxi Province, P.R. China</p> <p>Current position, grades: student of School of Information and Communication Engineering, North University of China, China. University studies: B.Sc. in Electronic information engineering from North University of China in China. M.Sc. from North University of China in China. Scientific interest: Acoustic signal processing, Experience: researching experience of 1 year, 1 scientific research projects.</p>

Motion analysis and simulation of a 12-Tetrahedral Walker Robot

Chun Li*, Hong Nie, Jinbao Chen

State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics
Nanjing, 210016, China

Received 6 January 2014, www.tsi.lv

Abstract

A novel robot mechanism-tetrahedral rolling robot is introduced in the paper. The robot comprises of 26 extension struts and 9 nodes. When the COG of tetrahedron exceeds the stability region, the robot will roll. The structure of the 12-TET robot is described. Designing method of the robot is given, and it is proved correct and feasible through simulation. Kinematic models in different motion phases are analysed in the paper, and the rolling critical condition is formulated. The effectiveness of the method is testified through simulation. The study of the paper will provide important reference for the dynamic analysis, optimization design and control of the tetrahedral rolling robot.

Keywords: Variable tetrahedron robot, Gait planning, Motion analysis

1 Introduction

A cutting-edge technology is Tetrahedral Robotics (TR), which involves active trusses made up of extensible structural members (struts) and interconnections (nodes) arranged in a tetrahedral mesh [1]. This conception is originally presented by Hamlin and Sanderson. The movement of a tetrahedral robot is realized through length change of the truss, and it is a novel mechanism with multi-DOF and multi-loop [2]. The robot moves by changing the lengths of several struts, which changes the location of the centre of gravity (COG). If the centre of gravity (COG) of the robot exceeds the stable region, it may cause the robot to “tumble”. The number of tetrahedra, or TETs, is an indicator of the complexity of a TR robot and the behaviours and shapes it can take on [3].

Tetrahedral Walker Robots provide greater mobility than conventional wheeled or track robots. This includes the abilities to [4]:

- 1) Traverse terrain more rugged in terms of slope, roughness, and obstacle size.
- 2) Conform to virtually any terrain.
- 3) Avoid falling down or getting stuck permanently.
- 4) Be easily maintained due to its modularity.
- 5) Avoid failure through compensatory gait (limp).

Tetrahedral Walker Robots architecture is being developed at NASA Goddard Space Flight Center (GSFC) for Lunar or Martian exploration [5]. The current line of work at GSFC on TR began in 2004 with the development of the 1-TET, a one-tetrahedron reconfigurable truss that demonstrated basic movement

and control of a determined system [6]. The second-generation 12-TET was deficient in a number of ways because of its weight. The third-generation TR 12-TET is so complex that it cannot move, and now the control method is in research [7, 8].

2 Mechanism design of tetrahedral walker robot

The mechanism sketch map of a variable 12-tetrahedral robot is shown in Figure 1.

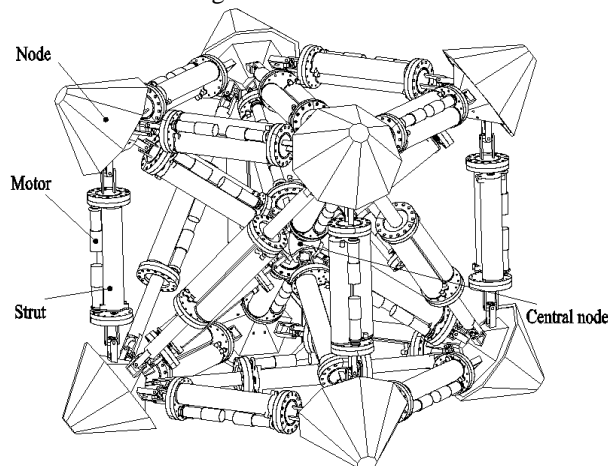


FIGURE 1 The configuration of a variable 12-tetrahedral robot

This 12-TET robot has an interior node for a payload and more continuous motion. The tetrahedral framework acts as a simple skeletal muscular structure [5]. The 12-TET is a mechanism with 9 nodes, and 26 struts which can look a lot like a box with a central node connected to each corner by a strut. There is a strut along each edge of the box and a diagonal strut on each face.

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

The geometry of the strut is a back-to-back double sided geometry expanding in two directions, improving the extension ratio with fewer segments and reducing crowding at the nodes. For actuation we have developed a

system of nested screws Figure 2 within an exoskeleton Figure 3, which has a greater extension ratio than we believe has previously been achieved.

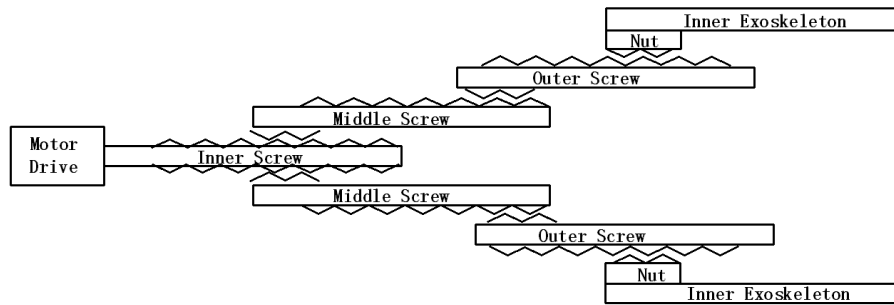


FIGURE 2 Screw design

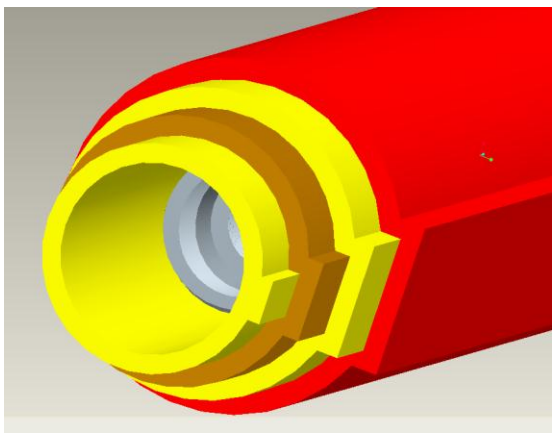


FIGURE 3 Exoskeleton design

Exoskeleton cannot spin, which keeps the last nut from spinning, which makes the makes the system progress linearly when the motor spins the inner screw.

3 Kinematic models analysis

The variable 12-tetrahedral robot’s movement is defined from the base configuration of the 12-TET shown in Figure 4a. The structure has 4 nodes that connect 7 struts (node 1, 3, 5 and 7), 4 nodes that connect 4 struts (node 2, 4, 6 and 8) and a centre node (node C)

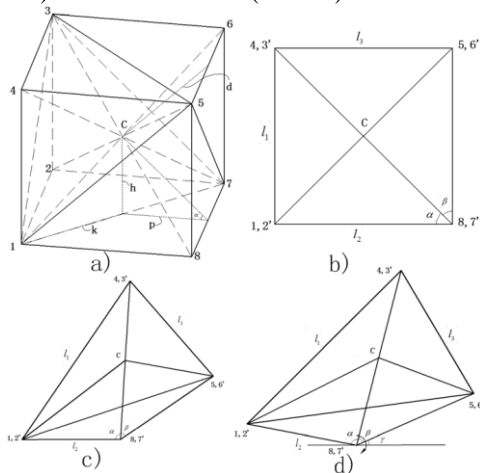


FIGURE 4 The base configuration of a variable 12-tetrahedral robot

3.1 TUMBLING MOVEMENT

We assume that the mass of the struts are negligible compared to that of the nodes, and each node’s mass of the variable 12-TET robot is equable. The variable 12-TET robot has different kinds of motion pattern, and one pattern that included two movements is chosen to analyse: Tumbling occurs over one of the four sides of the base (Square 1278). Supposing that robot would topple over the 78 strut. In the process of movement, node 3, 4, 7, 8 and node C is always located in the same Plane. For the first movement, the struts between Node a and Node b (L_{ab}) will be extending with the following length constraints (L is the initial length for the struts):

$$L_{14} = L_{23} = l_1, L_{13} = \sqrt{L^2 + l_1^2}, L_{18} = L_{27} = l_2, \\ L_{17} = \sqrt{L^2 + l_2^2}, L_{45} = L_{36} = l_3, L_{35} = \sqrt{L^2 + l_3^2}, \\ L_{1C} = L_{2C} = \sqrt{h^2 + k^2}, L_{5C} = L_{6C} = \sqrt{d^2 + \frac{L^2}{2}}, \\ L_{15} = \sqrt{L^2 + l_2^2 - 2Ll_2 \cos(\alpha + \beta)}, \\ \alpha = \arccos\left(\frac{2L^2 + l_2^2 - l_1^2}{2\sqrt{2}Ll_2}\right), \beta = \arccos\left(\frac{3L^2 - l_3^2}{2\sqrt{2}L^2}\right), \quad (1)$$

$$h = \frac{\sqrt{2}}{2} L \sin \alpha, p = \frac{\sqrt{2}}{2} L \cos \alpha, \\ k = \sqrt{\left(\frac{L}{2}\right)^2 + (L - p)^2}, d = \sqrt{\frac{3}{2}L^2 - \sqrt{2}L^2 \cos \beta}.$$

The rest of the struts do not change their lengths.

Now we consider the geometry of the robot and find the conditions under which the robot tumbles. We will show that tumbling occurs at certain ratios of the sides of the robot base. We should therefore change its dimensions accordingly. Once the torques about the 78 strut equals each other, the robot is going to topple. α is

the angle that rectangle 4378 relative to its initial position. The toppling critical condition is:

$$2l_2 + 2\sqrt{2}l_o \cos \alpha + \frac{\sqrt{2}}{2}l_0 \cos \alpha \geq -2l_0 \cos(\alpha + \beta),$$

$$\cos(\alpha + \beta) \geq \frac{5l_1^2 - 13l_2^2 - 10L^2}{8Ll_2} \quad (2)$$

3.2 KINEMATIC MODEL ANALYSIS BEFORE TOPPLING

3.2.1 Jacobi [J] and Hessian [H]

According to the mechanism characteristic and motion pattern of the variable 12-tetrahedral robot, the coordinates are established as Figure 5.

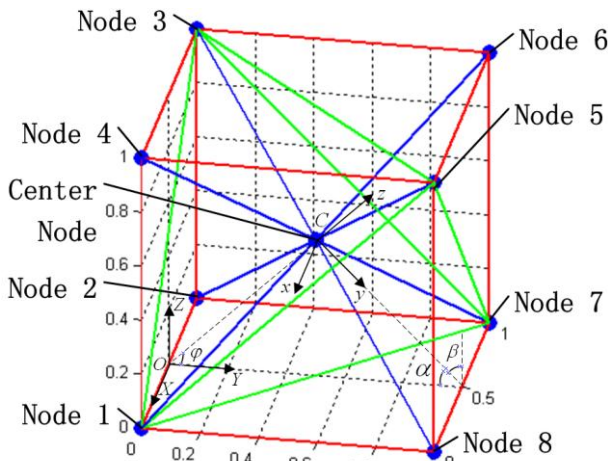


FIGURE 5 Coordinates of the variable 12-tetrahedral robot

Inertial frame (O-XYZ) and moving coordinates (C-xyz) is shown. Mid-point (O) of L_{12} for origin of inertial frame (O-XYZ), Axis X is perpendicular with plane 1234, Axis Y is superposed on line 12, and Axis Z can be deduced by the right hand rule. Centre node (C) for origin of moving coordinates (C-xyz), Axis z is perpendicular with $\Delta 8C7$, Axis x is superposed on midline of $\Delta 8C7$, and Axis y can be deduced by the right hand rule. Thus it can be seen, angle α between z and Z is the dihedral angle between plane 8C7 and plane 1278.

An arbitrary vector (R') in the moving coordinates (C-xyz) transformation to inertial frame (O-XYZ) given by:

$$R = [T]R' + PABC, \quad (3)$$

wherein

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix}, \quad (4)$$

$$P = [0 \ Y_c \ Z_c]^T, \quad (5)$$

in which $Y_c = l_2 - \frac{\sqrt{2}}{2}L \cos \alpha$ and $Z_c = \frac{\sqrt{2}}{2}L \sin \alpha$.

Supposing that the position and the pose of component i 's centre of the variable 12-tetrahedral robot is expressed as U_i, Φ_i , and U_i, Φ_i is the non-linear function of joint variable $l, l = (l_1, l_2, l_3)^T$. The equation is

$$U_i = f_1(l_1, l_2, l_3) \quad (6)$$

$$\Phi_i = f_2(l_1, l_2, l_3) \quad (7)$$

The first differential coefficient of them are the velocity v_i and angle velocity ω_i of component i . The equation is

$$v_i = \dot{U}_i = \begin{bmatrix} \frac{\partial f_1}{\partial l_1} & \frac{\partial f_1}{\partial l_2} & \frac{\partial f_1}{\partial l_3} \end{bmatrix} \begin{bmatrix} \dot{l}_1 & \dot{l}_2 & \dot{l}_3 \end{bmatrix}^T = J_{T,i} \dot{l}, \quad (8)$$

$$\omega_i = \dot{\Phi}_i = \begin{bmatrix} \frac{\partial f_2}{\partial l_1} & \frac{\partial f_2}{\partial l_2} & \frac{\partial f_2}{\partial l_3} \end{bmatrix} \begin{bmatrix} \dot{l}_1 & \dot{l}_2 & \dot{l}_3 \end{bmatrix}^T = J_{R,i} \dot{l}. \quad (9)$$

The second differential coefficient of them are the acceleration a_i and ε_i of component i . The equation is

$$a_i = \ddot{U}_i = \dot{J}_{T,i} \dot{l} + J_{T,i} \ddot{l} = H_{T,i} \dot{l} + J_{T,i} \ddot{l}, \quad (10)$$

$$\varepsilon_i = \ddot{\Phi}_i = \dot{J}_{R,i} \dot{l} + J_{R,i} \ddot{l} = H_{R,i} \dot{l} + J_{R,i} \ddot{l}. \quad (11)$$

$J_{T,i}, H_{T,i}$ are the translational Jacobi matrix and Hessian matrix, $J_{R,i}, H_{R,i}$ are the rotational matrix.

3.2.2 Jacobi [J] and Hessian [H] of nodes

Take centre node (C) for example, the position and the pose of C's centroid is:

$$U_C = P = \begin{bmatrix} 0 \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} 0 \\ l_2 - \frac{\sqrt{2}}{2} L \cos \alpha \\ \frac{\sqrt{2}}{2} L \sin \alpha \end{bmatrix}, \quad (12)$$

$$\Phi_C = \begin{bmatrix} \varphi \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \arctan\left(\frac{\sqrt{2} L \sin \alpha}{2l_2 - \sqrt{2} L \cos \alpha}\right) \\ 0 \\ 0 \end{bmatrix}, \quad (13)$$

where $\alpha = \arccos\left(\frac{2L^2 + l_2^2 - l_1^2}{2\sqrt{2}Ll_2}\right)$.

According to formula (8) ~ (11), the velocity v_C and angle velocity ω_C of node C is:

$$v_C = \dot{U}_C = J_{T,C} \dot{l} = J_{T,C} (l_1, l_2, l_3)^T, \quad (14)$$

$$\omega_C = \dot{\Phi}_C = J_{R,C} \dot{l} = J_{R,C} (l_1, l_2, l_3)^T \quad (15)$$

the acceleration a_C and ε_C

$$a_C = \ddot{U}_C = \dot{J}_{T,C} \dot{l} + J_{T,C} \ddot{l} = H_{T,C} \dot{l} + J_{T,C} \ddot{l}, \quad (16)$$

$$\varepsilon_C = \ddot{\Phi}_C = \dot{J}_{R,C} \dot{l} + J_{R,C} \ddot{l} = H_{R,C} \dot{l} + J_{R,C} \ddot{l}. \quad (17)$$

Similarly, we can obtain other nodes' J, H and velocity, acceleration.

3.2.3 Jacobi [J] and Hessian [H] of struts

The geometry of the struts is a back-to-back double sided geometry expanding in two directions, and the centroid in the geometric centre of struts. Take strut C5 for example; the position and the pose of strut C5's centroid in the moving coordinate is:

$$U_{C5} = \begin{bmatrix} \frac{x_5}{2} \\ \frac{y_5}{2} \\ \frac{z_5}{2} \end{bmatrix}, \quad \Phi_{C5} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix}, \quad (18)$$

where $\delta = \arccos\left(\frac{l_3^2 - L^2}{2Ll_3}\right)$, $x_5 = \frac{L}{2}$, $y_5 = \frac{l_3^2 - L^2}{2\sqrt{2}L}$, and $z_5 = \frac{\sqrt{4l_3^2L - l_3^4 - L^4}}{2\sqrt{2}L}$.

According to formula (3), the position and the pose of strut C5's centroid in the inertial frame (O-XYZ) is:

$$U_{C5} = [T]U'_{C5} + P, \quad \Phi_{C5} = [T]\Phi'_{C5} + P \quad (19)$$

The velocity v_{C5} and angle velocity ω_{C5} of strut C5 is:

$$v_{C5} = \dot{U}_{C5} = J_{T,C5} (l_1, l_2, l_3)^T, \quad (20)$$

$$\omega_{C5} = \dot{\Phi}_{C5} = J_{R,C5} (l_1, l_2, l_3)^T. \quad (21)$$

The acceleration a_{C5} and ε_{C5}

$$a_{C5} = \ddot{U}_{C5} = \dot{J}_{T,C5} \dot{l} + J_{T,C5} \ddot{l} = H_{T,C5} \dot{l} + J_{T,C5} \ddot{l} \quad (22)$$

$$\varepsilon_{C5} = \ddot{\Phi}_{C5} = \dot{J}_{R,C5} \dot{l} + J_{R,C5} \ddot{l} = H_{R,C5} \dot{l} + J_{R,C5} \ddot{l} \quad (23)$$

Similarly, we can obtain other struts' J, H and velocity, acceleration.

3.3 FALLING PHASE

The process of beginning the fall of the 12-TET robot is shown in fig.2. Supposing that the vector of strut 87 is:

$$\vec{u}_{87} = (u_x, u_y, u_z)^T. \quad (24)$$

The falling path of robot is expressed

$$\vec{v} = [R_{\beta,\gamma}] [R_{-\gamma,x}] [R_{\phi,z}] [R_{\gamma,x}] [R_{-\beta,\gamma}] \vec{v}_0 = [R_{\phi,u}] \vec{v}_0 \quad (25)$$

Rotational matrix

$$[R_{\phi,u}] = \begin{bmatrix} u_x^2 V \phi + \cos \phi & u_x u_y V \phi - u_z \sin \phi & u_x u_z V \phi + u_y \sin \phi \\ u_x u_y V \phi + u_z \sin \phi & u_y^2 V \phi + \cos \phi & u_y u_z V \phi - u_x \sin \phi \\ u_x u_z V \phi - u_y \sin \phi & u_y u_z V \phi + u_x \sin \phi & u_z^2 V \phi + \cos \phi \end{bmatrix} \quad (26)$$

where $V \phi = 1 - \cos \phi$, $\sin \gamma = u_y$, $\sin \beta = \frac{u_x}{\sqrt{u_x^2 + u_z^2}}$,

$\cos \gamma = \sqrt{u_x^2 + u_z^2}$, $\cos \beta = \frac{u_z}{\sqrt{u_x^2 + u_y^2}}$, $\cos \gamma \sin \beta = u_x$,

and $\cos \gamma \cos \beta = u_z$, ϕ is the roll angle of the robot.

Different walking paths can be planned through choosing different roll axis. Trajectory of the variable 12-

tetrahedral robot is shown in Figure 6. It is seen that the tetrahedral rolling robot could change its walking direction anytime, so it has great agility

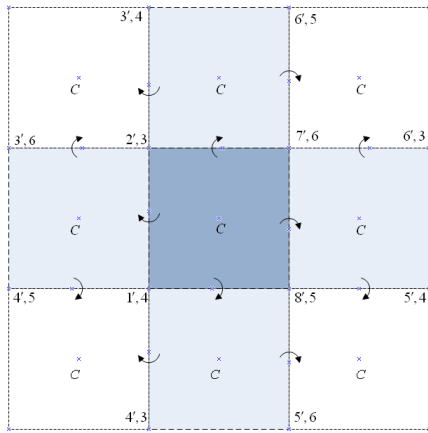


FIGURE 5 Trajectory of the variable 12-tetrahedral robot

4 Simulation

The simulation model is built for validating the effectiveness of the kinematic analysis method. The

parameter of the model is as follows: original length of strut $L=1000\text{mm}$.

We simulate the theoretical relations within MATLAB shown in Figure 6. This movement (Figure 7 a, b, c) moves the COG of the structure outside of the base and makes the robot tumble (Figure 7 d). The second movement (Figure 7 e, f) that completes the motion pattern involves the contraction of the same struts involved in the first movement. The contraction of the struts follows the same constraints of the first movement as well. Repeating this sequence of movement makes the robot move in a tumbling way.

The Adams simulation 12-TET robot is shown in Figure 8. It shows that the robot can realize the rolling motion according to the planned path, and then relapse. It indicates that the design method is feasible. The next movement can choose different roll axis and motion parameter; including direction, step length, and height. These can all be set neatly.

In the paper, the velocity of the extension strut is slow, and the dynamic effect is small. When the velocity increases, the dynamic effect will become greater. The next work will study this problem deeper.

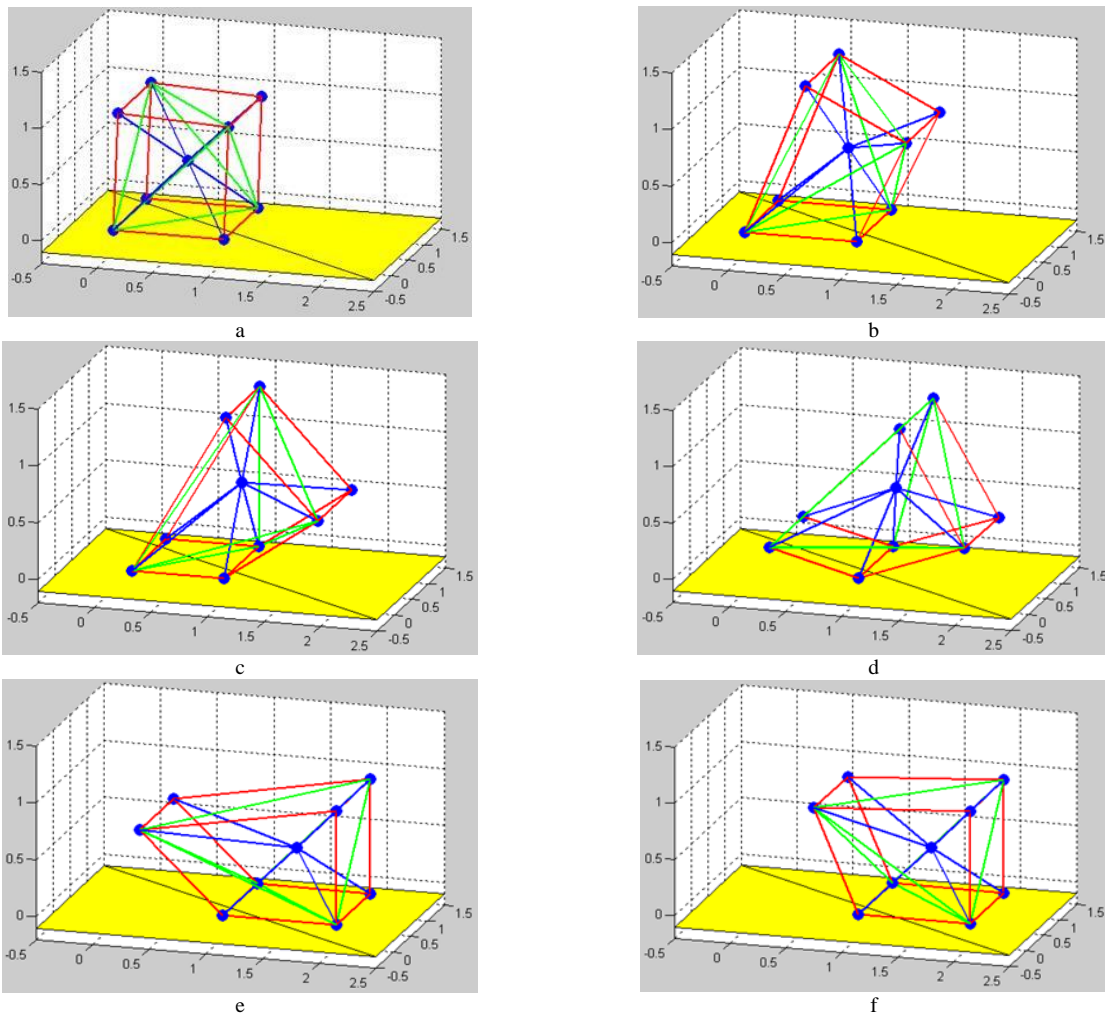


FIGURE 7 Rolling gait

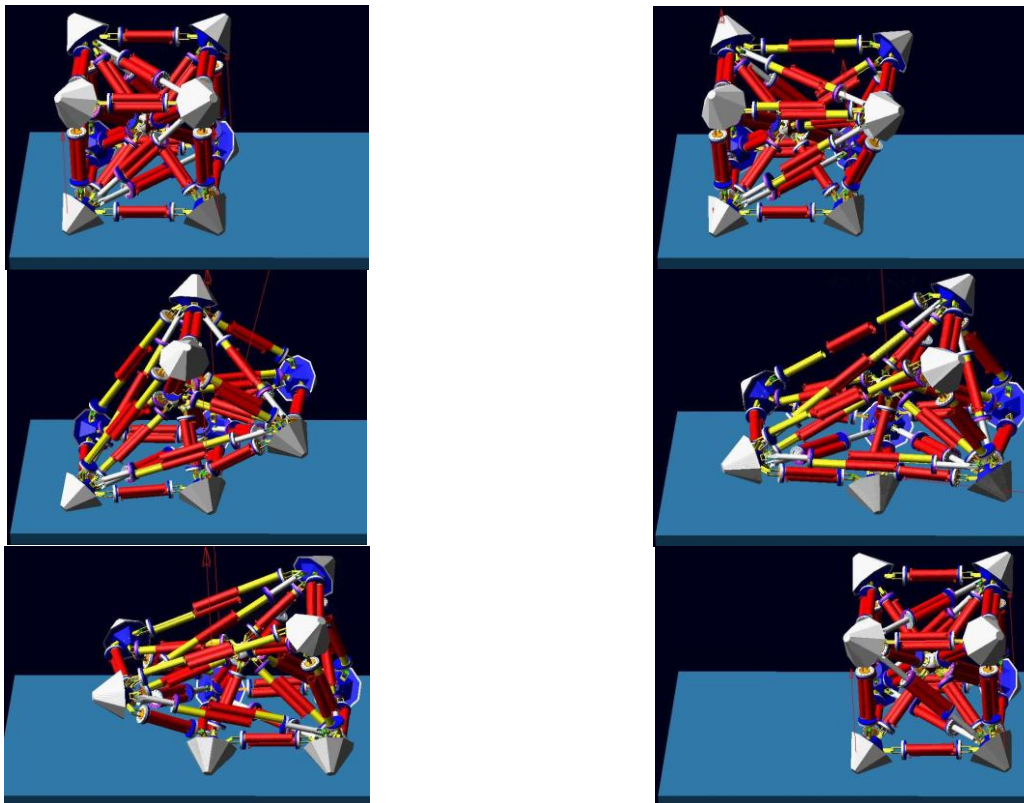


FIGURE 8 The Adams simulation 12-TET robot

References

[1] Curtis S, Brandt M, Bowers G 2007 Tetrahedral robotics for space exploration *IEEE Aerospace and Electronic Systems Magazine* 22(6) 22-30

[2] Miguel Abrahantes, Aaron Silver, Luke Wendt 2007 Gait Design and Modeling of a 12-Tetrahedron Walker Robot *39th Southeastern Symposium on System Theory* 10(2) 39-43

[3] Maziar Izadi, M J Mahjoob, Mohammad Soheilypour 2010 A Motion Planning for Toppling-Motion of a TET Walker *The 2nd International Conference on Computer and Automation Engineering* 8(4) 34-9

[4] Abrahantes M, Smits C 2012 Implementation and control of a reconfigurable 8-Tetrahedral robot *Electro/Information Technology (EIT), 2012 IEEE International Conference on* 6-8 May 2012 1-5

[5] Wan Ding, Sung-Chan Kim, Yan-An Yao 2012 A pneumatic cylinder driving polyhedron mobile mechanism *Frontiers of Mechanical Engineering* 7(1) 55-65

[6] M Abrahantes, D Littio, A Silver, L Wendt 2008 Modeling and Gait Design of a 4-Tetrahedron Walker Robot *System Theory, 2008. SSST 2008. 40th Southeastern Symposium on* 16-18 March 2008 269-73

[7] P E Clark, M L Rilee, S A Curtis 2004 BEES for ANTS: Space Mission Application for the Autonomous Nano Technology Swarm *AIAA 1st Intelligent Systems Technical Conference* 20-22 September 2004 1-12

[8] Abrahantes, M, Nelson L, Doorn P 2010 Modeling and Gait Design of a 6-Tetrahedron Walker Robot *System Theory (SSST) 2010 42nd Southeastern Symposium on* 5 248-52

[9] Zhang Lige, Bi Shusheng, Peng Zhaoqin 2011 Motion analysis and simulation of tetrahedral rolling robot *Journal of Beijing University of Aeronautics and Astronautics* 37(4) 415-20

Authors	
	<p>Chun Li, born in February, 1986, Nanjing, Jiangsu Province, P.R. China</p> <p>Current position, grades: Ph.D candidate of School of Machinery theory and design, Nanjing University of Aeronautics and Astronautics, China.</p> <p>University studies: B.Sc. in Aircraft Design and Engineering from Northwestern Polytechnical University in China (2006-2010).</p> <p>Scientific interest: a 12-tetrahedral rolling robot</p> <p>Publications: 3 patents.</p> <p>Experience: 1 scientific research project.</p>
	<p>Hong Nie, born in January, 1960, Anhui Province, P.R. China</p> <p>Current position, grades: Professor of School of Aircraft Design, Nanjing University of Aeronautics and Astronautics, China.</p> <p>University studies: Ph.D. in Aircraft Design from Northwestern Polytechnical University in China.</p> <p>Scientific interest: rolling robot, gear design</p> <p>Publications: 80papers, 30 patents.</p> <p>Experience: teaching experience of 23 years, 10 scientific research projects.</p>
	<p>Jinbao Chen, born in June, 1980, Nanjing, Jiangsu Province, P.R. China</p> <p>Current position, grades: Associate Professor of School of Nanjing University of Aeronautics and Astronautics, China.</p> <p>University studies: Ph.D. in Aircraft Design from Nanjing University of Aeronautics and Astronautics in China (2005-2008).</p> <p>Scientific interest: gear design, soft landing of lunar</p> <p>Publications: 20papers.</p> <p>Experience: teaching experience of 4 years, 6 scientific research projects.</p>

Numerical and analytical solution of stresses on a box-type lining structure under the effect of ground fracture

W Huang^{1, 2}, D Y Liu^{1, 2, 3*}, H F Jiang^{1, 2}, F Y Liu^{1, 2}

¹College of Civil Engineering, Chongqing University, Chongqing, China, 400045

²Key Laboratory of New Technology for Construction of Cities in Mountain Area (Chongqing University) Ministry of Education, Chongqing, China, 400045

³School of Civil Engineering and Architecture, Chongqing University of Science & Technology Chongqing, Chongqing, China, 401331

Received 5 April 2014, www.tsi.lv

Abstract

This study attempts to study the stress mechanism of a box-type lining structure during its inclined penetration of ground fissure and calculate the normal stress and shearing stress of the structure section. Based on thin-wall structure theory combined with the stress boundary conditions of the physical model of the box-type lining structure, we derived the analytical solution of the normal stress and shearing stress of the physical model. The stress analytical solution indicates that the damage of the footwall of the ground fissure is more serious than that of the hanging wall, which could match the physical model experiment. The effectiveness and accuracy of the analytical solution of the normal stress and shearing stress of the section were verified using finite element software to establish the mechanical model of the box-type lining structure. The results of the numerical model were compared with the analytical solution results.

Keywords: Box-type structure, Lining structure, Ground fissure, Analytical solution

1 Introduction

Xi'an City, China is one of the cities with the most intensely developed ground fractures in the world. When the box-type lining structure of the subway in Xi'an penetrates the active ground fracture, the tunnel structure becomes subject to the bending stress and shearing stress under the pressure of the surrounding rocks. This condition is caused by the faulting of the hanging wall and the heading wall of the ground fracture. The normal rate of the vertical movement of the ground fracture in the area is 5 to 35 mm/a , whereas the maximum rate is 55.06 mm/a [1]. The intense ground fissure movement threatens the subway operation. At the same time, the current designing codes do not cover the stress design of the section of the integral box-type lining structure, and no code could be followed for the designing of the project [2].

With the development of underground space in the 20th century, tunnelling technology has witnessed rapid development. Rich experiences have been accumulated in terms of ground fissure research [3], stress analysis of tunnel structures [4-9], and the construction of subway tunnels [10]. Pavlovic et al. [11] conducted a numerical simulation on the bending and torsion performance of a thin-wall box-type column to analyse the influence of the initial fault and residual stress on the bending stability. Park et al. [12] conducted an on-site survey and material

test to analyse the damage caused by a fire accident to the box-type subway tunnel in Daegu City, South Korea and propose a remedy plan. Stirbys et al. [13] used structural processing measures, such as expanding section and the combination of solid support and flexible support, to resolve the issues encountered during the penetration of the tunnel through the faults.

Although other researchers [14, 15] have exposed the cause of the formation of the ground fissures in Xi'an and the analytical solution of the torsional deformation [16], studies on the analytical solution of the bending and shearing stresses on the integral box-type lining structure when penetrating through active ground fissures are rare.

This study focuses on the disaster-causing mechanism of active ground fractures over box-type subway tunnels and decreasing the disastrous effect of the ground fractures on the subway operation. We have deduced the analytical solution of the bending and shearing stresses of an integral box-type lining structure during the inclined penetration through active ground fractures. We have also established a finite element model to analyse the bending and shearing deformations of the box-type structure to provide theoretical and model references for the subway structure designing in ground fracture areas. The results can help to enrich the knowledge on the stress influence of subway structures crossing ground fracture sections.

* Corresponding author email: liudy@cuq.edu.cn

2 Model Stress Analysis

2.1 MODEL OVERVIEW

The foundation settlement test platform is used to establish the physical model of the integral box-type lining structure penetrating ground fractures according to a ratio of similitude of 1:5. The disturbed soil (silty clay)

of the actual stratum in Xi'an is used for the test. The ground fracture is imitated by filling with silty-fine sand with a dip angle of 85°, which intersects with the axes of the structure at 30°. The major physico-mechanical indexes could basically meet the similarity relation. Figure 1 shows the section of the soil layers. Reference 1 lists the detailed settings of the physical model.

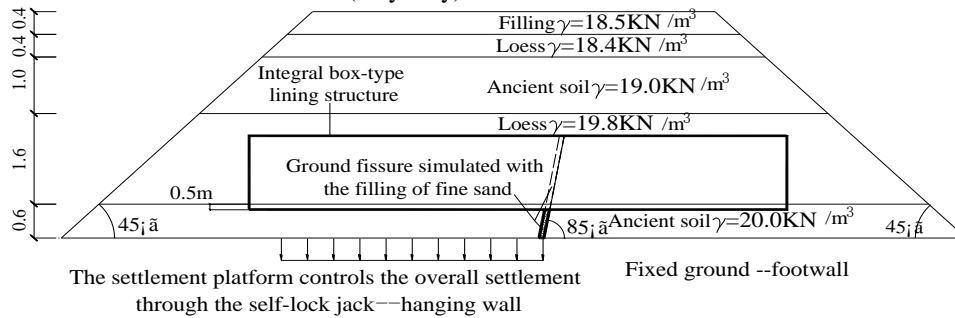


FIGURE 1 Physical model of integral box-type lining structure

2.2 BOUNDARY CONDITIONS OF MODEL STRESS

The section centreline dimension of the integral box-type lining structure was $d_1 \times d_2 = 1182\text{mm} \times 1090\text{mm}$, the wall thickness was $d = 120\text{mm}$, the structural length of the physical model was 10 m, and the ground fracture was 6 m away from the left end of the structure. During the test, the heading wall of the ground fracture was fixed, and the settlement of the ground fracture was simulated with the operation rate and movement of the self-lock jack installed on the hanging wall of the ground fracture. With the relative lowering of the hanging wall of the ground fracture, an empty space was formed at the bottom of the hanging wall. According to the available test results, the basal contact pressure of the maximum earth pressure cell was zero within 0 to 0.5 m from the heading wall to the ground fracture because the lining structure is subject to bending deformation. With the increase in the settlement, the contact pressure at the bottom of the hanging wall on the left end surface of the lining structure gradually decreased to zero. This condition indicates that the empty space at the bottom of

the lining structure has expanded to the entire bottom of the heading wall with the increase in the settlement.

According to the observations of the physical model test, the boundary conditions of the lining structure could be concluded as follows. The top of the box-type lining structure was under the pressure load from the uniformly distributed soil P. This condition subjected the structure to the bending and shearing damages at the empty space of the hanging wall and the local area near the ground fracture at the heading wall Figure 2. The left end surface of the lining structure was buried in the fourth and fifth filling layers Figure 1. Its movement in the axial direction X was smaller than the movement of the empty space along the vertical direction Z. Thus, the left end surface of the lining structure was considered as the directional supporting base Figure 2. The right end surface of the lining structure was on the heading wall that was relatively fixed, which was also buried in the fourth and fifth layers of the soil Figure 1. Thus, the right end surface of the lining structure could be considered as the fixed supporting base Figure 2.

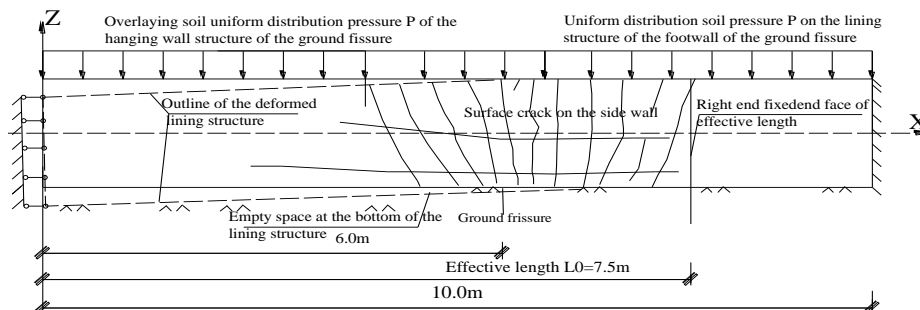


FIGURE 2 Stress Diagram of Integral Box-Type Lining Structure

3 Analytical Solution of the Bending and Shearing Stresses of the Box-Type Lining Structure

3.1 FUNDAMENTAL ASSUMPTIONS

(1) Given the wall thickness ($d = 120\text{mm}$) is smaller than the width of the section and the length of the structure, the shearing stress flow generated from the

shearing deformation is assumed to be uniformly distributed along the wall thickness.

(2) The plane cross-section assumption is used in calculating the bending deformation of the integral box-type lining structure.

(3) In the physical model, the ground fracture is 6 m away from the left end surface of the structure, whereas the overall length of the integral box-type lining structure is 10 m. Given that the bottom of the box-type lining structure was subject the maximum bending deformation within 0 to 0.5 m of the heading wall of the ground fracture, the soil body at the bottom collapsed to move downward and form an empty space. Thus, the scope of the empty space was $x=0\sim 6.5\text{m}$, whereas the right end of the structure's effective length was fixed within

$x=6.5\sim 10.0\text{m}$. According to the test results, the effective length of the box-type lining structure is assumed to be $L_0=7.5\text{m}$ Figure 2.

3.2 BENDING NORMAL STRESS OF THE INTEGRAL BOX-TYPE LINING STRUCTURE

According to the fundamental assumptions, the integral box-type lining structure was subject to the bending stress over the effective length of $L_0=7.5\text{m}$ under the overlying uniform distribution soil pressure P . The cross-sections that are vertical to the axes of the tunnel Figure 3a were still plane after the structure bended, and were vertical to the axes of the deformed lining structure.

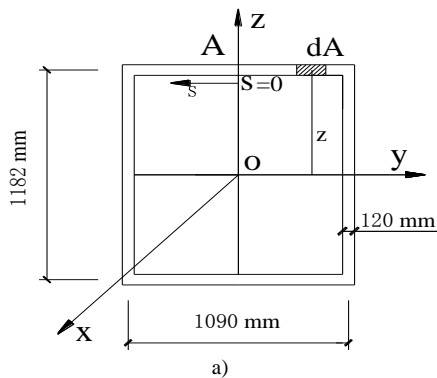


FIGURE 3 Cross Section of the Integral Box-Type Lining Structure

Figure 3a) shows a cross section of the box-type lining structure. $o-xyz$ is a rectangular coordinate system crossing the section centroid of the section, whereas $o-x$ is the axis of the lining structure. The intersection point of axis $o-z$ and the middle surface of the structure is A , s is the curve coordinate of the outline of the middle surface of the structure, and the anti-clockwise direction is the positive direction. A is the initial point of the curve coordinate s . The wall thickness was $d = 120\text{mm}$.

According to the plane cross-section assumption, the normal strain at any arbitrary point on the section is as follows:

$$\varepsilon_x = \frac{z}{\rho}, \tag{1}$$

where ρ is the curvature radius of the neutral layer of the section, and z is the vertical distance between any point and the neutral layer.

The normal stress of the section is as follows:

$$\sigma_x = \frac{Ez}{\rho}. \tag{2}$$

According to the torque equilibrium equation surrounding axis y :

$$M_y = \int_A \sigma_x z dA. \tag{3}$$

Placing (2) into (3) achieves the following:

$$M_y = \int_A \frac{Ez^2}{\rho} dA = \frac{E}{\rho} I_y. \tag{4}$$

Placing (4) into (2) achieves the following:

$$\sigma_x = \frac{M_y}{I_y} z \tag{5}$$

where M_y is the bending moment surrounding axis y on section x , and I_y is the moment of inertia surrounding axis y .

3.3 SHEARING STRESS OF THE INTEGRAL BOX-TYPE LINING STRUCTURE

Considering that, the integral box-type lining structure is relatively fixed at the footwall, its part at the hanging wall moves downward. Each section of the lining structure is subject to the shearing stress caused by the overlying soil pressure P . Suppose a unit $ds \times dx$ with a thickness of $d = 120\text{mm}$ at the initial point of the coordinate s Figure 4.

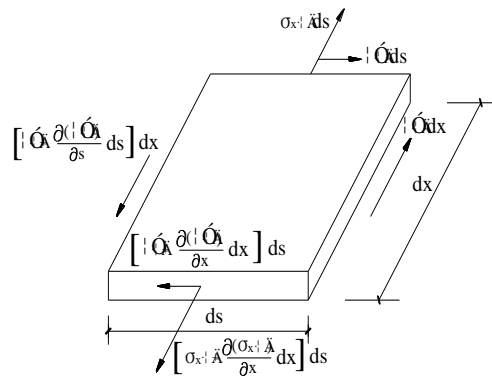


FIGURE 4 Axial Projection Drawing of the Side-Wall Stress of the Structure

Given that $\sum x = 0$:

$$\frac{\partial(\sigma_x \delta)}{\partial x} + \frac{\partial(\tau \delta)}{\partial s} = 0. \tag{6}$$

Considering that the wall thickness $d = 120\text{mm}$ is constant to irrelevant to x , thus:

$$\tau \delta = -\int_0^s \delta \frac{\partial \sigma_x}{\partial x} ds + C_1(x). \tag{7}$$

Placing (5) and the shear $Q_z = \partial M / \partial x$ at the direction z into (7), and assuming the shearing stress flow $q = \tau \delta$, then:

$$q = q_0 + q_A = -\left(\frac{Q_z}{I_y} S_y\right) + (\tau \delta)_A, \tag{8}$$

where $q_A = (\tau \delta)_A$ is the shearing stress flow of the initial point A of the curvilinear coordinates of the section.

We have introduced the continuous conditions [17, 18] $\oint u(x, s) ds = 0$ of the warping movement of the section of the box-type lining structure. These conditions are set to $u(x, s)$ as the axial warping displacement of the section x coordinate to calculate the axial warping displacement of point s in the section through the shear deformation $du / ds = \gamma = q / (G\delta)$.

$$u(x, s) = \int_0^s \frac{q}{G\delta} ds + u_A. \tag{9}$$

The integral initial point of formula (9) is point A . From the initial point A , the line integral is conducted along the outline of the section to A . The relative axial displacement is zero. Thus:

$$u - u_A = \oint \frac{q}{G\delta} ds = 0. \tag{10}$$

Placing (8) into (10) achieves the following:

$$q_A = -\oint \frac{Q_z}{\delta I_y} S_y ds / \oint \frac{ds}{\delta}. \tag{11}$$

Placing (11) into (4) achieves the section shearing stress flow:

$$q = q_0 + q_A = \frac{Q_z}{I_y} (-S_y + \frac{\oint \frac{S_y}{\delta} ds}{\oint \frac{ds}{\delta}}), \tag{12}$$

where Q_z is the shear in the direction z on section x , S_y is moment of area of point s related to axis y , and $d = 120\text{mm}$ is the wall thickness of the section.

4 Analytical Solution of the Bending Normal Stress and Shearing Stress of the Integral Box-Type Lining Structure

4.1 SOLUTION TO THE BENDING NORMAL STRESS OF THE SECTION

According to the abovementioned boundary conditions and fundamental assumptions, the following factors are determined. The effective length of the integral box-type lining structure was $L_0 = 7.5\text{m}$. The left end surface of the structure is the directional supporting base. The right end surface is the fixed supporting base. The distance between the ground fracture and the left end surface was 6 m. The overlaying uniform distribution soil pressure was $p = 0.065\text{MPa}$, whereas the linear load along the axes of the lining structure was $q = 0.065 \times 1210 = 78.65\text{kN/m}$. The bending moment (which makes the tension on the lower part of the section and the pressure on the upper part of the section positive) and the shear (Table 1) of each $L/15$ section are calculated.

According to (1), $I_y = 0.125735089\text{m}^4$ and $z = 0.591\text{m}$. The maximum bending normal stress of each $L/15$ section of the box-type lining structure could be then calculated Table 1.

TABLE 1 Maximum Normal Stress and Shearing Stress of the Integral Box-Type Lining Structure Section

Section Location	Bending Moment M_y	Shearing Stress Q_z	Maximum Normal Stress of the Analytical Method	Maximum Normal Stress of the Numerical Method	Maximum Shearing Stress of the Analytical Method	Maximum Shearing Stress of the Numerical Method
	kN*m	kN	MPa	MPa	MPa	MPa
0	737.625	0.0	3.467	3.617	0.000	0.000
L/15	452.238	39.325	2.126	2.214	0.155	0.151
2L/15	186.794	78.650	0.878	0.954	0.311	0.319
3L/15	-58.988	117.975	-0.277	-0.203	0.466	0.487
4L/15	-285.106	157.300	-1.340	-1.584	0.621	0.654
5L/15	-491.563	196.625	-2.311	-2.547	0.777	0.822
6L/15	-678.356	235.950	-3.189	-3.406	0.932	0.990
7L/15	-845.4875	275.275	-3.974	-4.289	1.088	1.158
8L/15	-992.956	314.600	-4.667	-4.863	1.243	1.321
9L/15	-1120.763	353.925	-5.268	-5.487	1.398	1.494
10L/15	-1228.906	393.250	-5.776	-5.972	1.554	1.661
11L/15	-1317.388	432.575	-6.192	-6.411	1.709	1.829
12L/15	-1386.206	471.900	-6.516	-6.766	1.864	1.997
13L/15	-1435.363	511.225	-6.747	-6.951	2.020	2.165
14L/15	1464.856	550.550	-6.885	-7.214	2.175	2.332
L	1474.688	589.875	-6.932	-7.481	2.330	2.500

4.2 SOLUTION TO THE SHEARING STRESS OF THE SECTION

The linear load of the lining structure along the axes was $q=78.65kN/m$. The left end surface of the structure model was the directional supporting base, whereas the right end surface was the fixed supporting base. Thus, the shear Q_z on each $L/15$ section of the structure could be calculated. The maximum shearing stress of each $L/15$ section Table 1 could be calculated by placing the above shear Q_z into (12), which occurs halfway the height of the section.

5 Finite Element Model

The 3D finite element model of the integral box-type lining structure was established according to the severe of the soil layer shown in Figure 5. This model verifies the maximum normal stress and shearing stress of the section obtained through the analytical method.

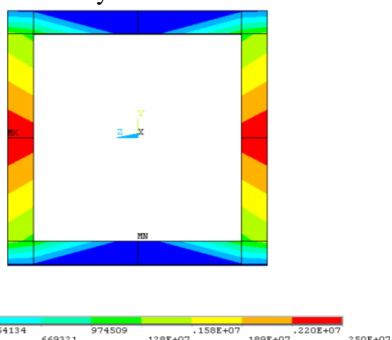


FIGURE 5 Diagram of the Shearing Stress on the Right End Section Surface

The BEAM188 structure unit to simulate the box-type lining structure, which is in contact with the interfaces set for the soil layers, was considered. The left end surface of the lining structure was set as the directional supporting base, whereas the right end surface was set as the fixed

supporting base. According to the actual settlement, the hanging wall of the ground fissure was gradually decreased to 20cm in the software. Table 1 shows the maximum normal stress and shearing stress of each section $L/15$ calculated at the final settlement of 20cm. Figure 5 shows the maximum normal stress and the shearing stress of the right end surface of the structure.

6 Conclusion

The analytical solution of the normal stress and shearing stress of the bending deformation of an integral box-type lining structure passing through ground fractures has been calculated. This solution is based on the physical model test and boundary conditions (i.e., the left end is the directional supporting base, whereas the right end is the fixed supporting base). The conclusions and suggestions of this study are as follows:

- (1) With the lowering of the hanging wall of the ground fracture and the formation of the empty space, the lower side of the section of the lining structure (0~2L/15) was subject to the tension. The upper side of the 3L/15~L section was subject to the tension. The maximum normal stress along the axes of the structure increased gradually. At the right end section, the normal stress reached a maximum value of 6.932MPa, which is significantly larger than the standard strength of extension of C30 concrete.
- (2) The maximum section shearing stress of the integral box-type lining structure is subject to linear increase along the axes and reaches the maximum value at the right end section. The maximum shearing stress is achieved at halfway the height of the section.
- (3) The maximum normal stress and shearing stress on the section of the heading wall of the ground fracture (12L/15) are larger than those of the hanging wall. During the bending of the structure, the ring-direction cracks

could be seen on the heading wall section first. The damage to the heading wall because of the bending is more serious than that to the hanging wall.

(4) The damage effect of the integral box-type lining structure is large when passing through ground fractures. Special deformation joints are set to release the stresses on the structure near the ground fractures. Water proof measures also need to be taken.

References

- [1] Ji Y S 2009 *Model test study on the interaction of subway tunneling and surrounding rock in ground fissure* Master Thesis of Chang'an University, China.
- [2] Huang W 2011 *Study on analytic solution of torsional deflection on subway tunnel structure inclined wear activities ground fissure* Master Thesis of Chang'an University, China.
- [3] Liu L, Chen L Q, Tang J X 2010 Present Situation and Future Prospects of Geologic Environment Issues in Mines in China *Disaster Advance* 3(4) 563-6
- [4] Stirbys A F, Radwanski Z R, Proctor R J 1999 Los Angeles metro rail project-geologic and geotechnical design and construction constraints *Engineering Geology* 51(3) 203-24
- [5] Pavlovic L, Froschmeier B, Kuhlmann U 2012 Finite element simulation of slender thin-walled box columns by implementing real initial conditions *Advances in Engineering Software* 44(1) 63-74
- [6] Ling D S, Guo H G, Wu J 2012 Research on seismic damage of metro station with centrifuge shaking table model test *Journal of Zhejiang University (Engineering Science)* 46(12) 2202-9
- [7] Skee A, Bjelanovic G, Jelenic G 2013 Glued timber-concrete beams-analytical and numerical models for assessment of composite action *Engineering Review* 33(1) 41-49
- [8] Wang S R, Chang M S 2012 Reliability Analysis of Lining Stability for Hydraulic Tunnel under Internal Water Pressure *Disaster Advance* 5(4) 166-70
- [9] Kultasov K A, Rysbekova A A 2012 Calculation of composite plates of variable thickness and fold in action radial force *Computer Modelling & New Technologies* 16(3) 54-57
- [10] Duan B F, Li L 2012 Grouting Reinforcement Technique for Subsurface Excavation of Group Metro Tunnels Construction in Complex Environment *Disaster Advance* 5(4) 1508-12
- [11] Pavlovic L, Kuhlmann U, Froschmeier B 2012 Finite element simulation of slender thin-walled box columns by implementing real initial conditions *Advances in Engineering Software* 44(1) 63-74
- [12] Park S Y, Oh H H, Shin Y S 2006 A case study on the fire damage of the underground and its repair works *Tunnelling and Underground Space Technology* 21(3) 328
- [13] Stirbys A F, Radwanski R Z, Proctor R J, Escandon R F 1999 Los Angeles metro rail project-geologic and geotechnical design and construction constraints *Engineering Geology* 51(3) 203-24
- [14] Peng J B, Hu Z P, Men Y M 2009 Test study of deformation and damage mechanism of horseshoeshaped tunnel crossing ground fissure with 40° *Chinese Journal of Rock Mechanics and Engineering* 28(11) 2258-65 (in Chinese)
- [15] Fan W, Deng L S, Peng J B 2008 Study of the physical model experiment of subway tunnel crossing the ground fissure belt *Chinese Journal of Rock Mechanics and Engineering* 27(9) 1917-23 (in Chinese)
- [16] Liu D Y, Huang W, Luo L J, Li D S, Zhao B Y 2013 Torsion deformation's analytic solution of box tunnel crossing ground fissure *Journal of Civil, Architectural & Environmental Engineering* 35(1) 1-6 (in Chinese)
- [17] Huang J Y 1998 *The torsion analysis of thin-walled structure - The curve beam and diagonal branch box beam* China Railway Publishing House: Beijing (in Chinese)
- [18] Zhang Y H 2008 Study on shear stress of thin-walled member with closed section under restrained torsion *Journal of Lanzhou Jiaotong University* 27(4) 1-3 (in Chinese)

Acknowledgements

This work was funded by the National Natural Science Foundation Funding Project (Grant Nos. 40602033 and 41302223) and the Chongqing City Fundamental and Advanced Research Projects (cstc2013jcyjA90004).

Authors	
	<p>Huang W, born in April, 1984, College of Civil Engineering, Chongqing University, Chongqing, P.R. China</p> <p>Current position, grades: Doctor of College of Civil Engineering, Chongqing University, China. University studies: BEng. in College of Civil Engineering from Chang'an University in China. ME Civil Engineering from Chang'an University in China. Scientific interest: Geotechnical Engineering, Underground Structure. Publications: 5 papers published in various journals. Experience: 5 scientific research projects.</p>
	<p>Liu D Y, born in October, 1959, College of Civil Engineering, Chongqing University, Chongqing, P.R. China</p> <p>Current position, grades: Professor of Civil Engineering, Chongqing University, China. University studies: BEng from Chengdu university of Science & Technology in China. D.E. from Chongqing University in China. Scientific interest: Geotechnical Engineering, Underground Structure Publications: 120 papers published in various journals. Experience: 20 scientific research projects.</p>
	<p>Jiang H F, born in December, 1982, College of Civil Engineering, Chongqing University, Chongqing, P.R. China</p> <p>Current position, grades: Doctor of College of Civil Engineering, Chongqing University, China. University studies: BEng in College of Civil Engineering from Chongqing Jiaotong University in China. ME Civil Engineering from Chongqing Jiaotong University in China. Scientific interest: Geotechnical Engineering. Publications: 6 papers published in various journals. Experience: 6 scientific research projects.</p>
	<p>Liu F Y, born in December, 1989, College of Civil Engineering, Chongqing University, Chongqing, P.R. China</p> <p>Current position, grades: the Master of College of Civil Engineering, Chongqing University, China. University studies: BEng in College of Civil Engineering from Chongqing University in China. Scientific interest: Geotechnical Engineering, Underground Structure Publications: 3 papers published in various journals. Experience: 2 scientific research projects.</p>

Numerical simulation on methane explosion propagation in a one-dimensional straight duct with porous metal materials

Zhenzhen Jia^{1, 2}, Feng Tao^{1, 2*}

¹School of Resources and Safety Engineering Central South University, Changsha City, Hunan Province, China, 410083

²School of Energy and Safety Engineering Hunan University of Science and Technology, Xiangtan City, Hunan Province, China, 411201

Received 6 October 2013, www.tsi.lv

Abstract

Based on the theoretical and experimental results of methane explosion propagation in porous metal materials, methane explosion propagation in a one-dimensional straight duct with different layers of porous metal materials is simulated by *Fluent* software. The layers and length of porous metal materials have influence on the flame propagation velocity and explosion shockwave overpressure. Only the propagation distance of methane explosion is beyond 5 times the diameter of the duct, the flame and explosion wave can be attenuated by the porous metal materials. Moreover, the more the layers are, the better the attenuation effect is. The numerical simulation results show that, during methane explosion propagation in porous metal materials, the attenuations of explosion wave overpressure and flame propagation velocity takes on synchronization and correspondence. Consequently, the porous metal materials can suppress methane explosion propagation. The process of methane explosion in the duct well reappears with numerical simulation, thus the model is established correctly and the numerical simulation is a good supplementary means of experiment.

Keywords: Methane Explosion, Porous Metal Materials, Flame Propagation Velocity, Explosion Shockwave Overpressure, Fluent Software

1 Introduction

Methane explosion is a serious disaster in mines, thus how to prevent and control methane explosion has always been a hot research and the focus of everyone's attention. Nevertheless, methane explosion prevention is mainly manifested in two aspects: how to attenuate explosion wave and how to extinguish flame. At present, during methane explosion propagation in porous metal materials, the theory on flame quenching mainly includes the thermal effect, the wall effect and the flame expansion principle [1, 2]. Moreover, when explosion shockwave and porous metal materials interact, the porous metal materials can attenuate explosion shockwave [3, 4].

Gvozdeva, Dupe, Vasil'ev, Teodoreczyk, Reddy et al. found that porous compressible materials can significantly attenuate the pressure wave intensity, and the absorption material installed on the pipe wall can effectively suppress the detonation wave [5-9]. Yu Jianliang et al. experimentally studied the relationship among three parameters (i.e. critical quenching velocity, critical quenching pressure and quenching parameter) and geometry parameter of the multi-layer wire mesh suppressive structure and obtained the empirical formula of the relationship [10]. Wang Zhencheng and Terushige Ogawa experimentally studied the suitable wire gauze parameter, and obtained the equation among the bound flame speed, wire gauze shape coefficient (d/W) and the number of the layer of wire gauze [2]. Fedorov and Fedorchenko studied the problem of shockwave

interaction with a layer of a porous material on a solid surface [11]. In addition, many scholars numerically simulated the explosion shockwave overpressure and flame propagation velocity during methane explosion in porous metal materials, obtained that porous metal materials have the actions of flame quenching and shockwave attenuation on methane explosion, and found that porous metal materials can suppress methane explosion propagation [12, 13].

According to the above literatures, the porous metal materials have suppression effects on methane explosion, but the influence of porous metal material parameters on methane explosion suppression cannot be determined. Based on methane explosion propagation characteristics in porous metal materials and *Computational Fluid Dynamics* theory, the mathematical model is established, the appropriate algorithm is selected, and the attenuation law of methane explosion propagation in a one-dimensional straight duct with different layers of porous metal materials (the number of layers is equal to 0, 1, 2, 3, 4, 5, respectively) is numerically simulated by *Fluent* software.

2 Geometric modelling and mesh generation

In the process of numerical simulation, the duct was simplified as a two-dimensional geometric modelling Figure 1. The length of the duct was 8.5 m, the diameter of duct section was 15 cm, the porous metal material was installed on the duct wall in the range of 4.5-7.5 m from

* Corresponding author e-mail: 66896105@qq.com

the left end, the ignition system was located on the left of the helical accelerating ring, the left end was close and the right end was open.

The computational grid was generated by *GAMBIT* with quadrilateral cells. In the axial direction, the mesh

was uniform, and its size was about 8 mm. In the cross section, the mesh near the wall was fine and its size was about 1 mm. And mesh size was about 5 mm in the centred zone. The mesh contains 48482 nodes, 46288 quadrilateral cells and 94769 faces.

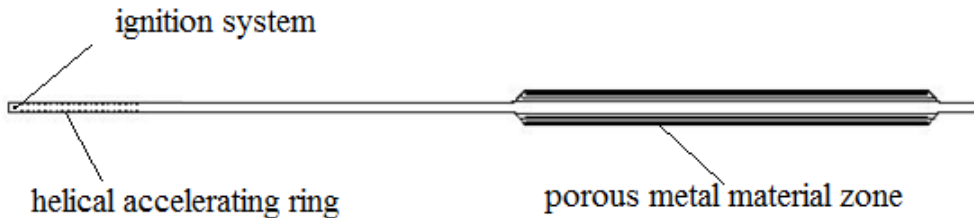


FIGURE 1 Geometric modelling

3 Numerical modelling

Before the numerical modelling was built, we hypothesized that: (1) the premixed methane-air mixture is considered as ideal methane, (2) the methane concentration in pre-mixed methane-air mixture is 9.5% (Concentration percentage value refers to percent by volume), (3) heat radiation and volume force can be ignored.

3.1 GOVERNING EQUATIONS

Methane explosion propagation in the duct is governed by the conservation laws for mass, momentum and energy. In the methane explosion process, if the combustion and supersonic flow need to be calculated, the density fluctuation cannot be ignored. Therefore, the differential control equations are obtained by the density weighted average concept proposed by Favre. They can be described in the following equations [14].

The continuity equation can be written as:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j) = 0. \tag{1}$$

The conservation equation for momentum can be written as:

$$\begin{aligned} \frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j) = & -\frac{\partial P}{\partial x_i} \\ & + \frac{\partial}{\partial x_j} \left[u_e \left(\frac{\partial u_j}{\partial x_i} \right) - \frac{2}{3} \left(\rho k + u_e \frac{\partial u_k}{\partial x_k} \right) \delta_{ij} \right]. \end{aligned} \tag{2}$$

The conservation equation for energy can be written as:

$$\frac{\partial(\rho h)}{\partial t} + \frac{\partial}{\partial x_j} \left(\rho u_j h - \frac{u_e}{\sigma_k} \frac{\partial h}{\partial x_j} \right) = \frac{dP}{dt} + S_h. \tag{3}$$

Composition equation can be written as:

$$\begin{aligned} \frac{\partial \rho Y_{fu}}{\partial t} + \frac{\partial}{\partial x_j} \left(\rho u_j Y_{fu} - \frac{u_e}{\sigma_{fu}} \frac{\partial Y_{fu}}{\partial x_j} \right) = R_{fu}, \\ \delta_{ij} = \begin{cases} 1(i = j) \\ 0(i \neq j) \end{cases}; u_e = u + u_t, \end{aligned} \tag{4}$$

where, t is time, ρ is the density of a material, P is pressure, h is total enthalpy, σ_k is effective Prandtl Number of turbulent kinetic energy diffusion, σ_e is the Prandtl Number of Pulsation kinetic energy dissipation rate ϵ . The subscript i, j, k are summation conventions, u_i is velocity component in the i direction, k is turbulent kinetic energy, Y_{fu} is fuel mass fraction, δ_{ij} is Kronecker- δ function, u viscosity coefficient of laminar flow, u_t is viscosity coefficient of turbulent flow, R_{fu} is chemical reaction rate, namely, combustion rate.

3.2 TURBULENCE MODEL

In the simulation, the RNG k- ϵ model was used to be the turbulence model. The RNG k ϵ model was derived using a statistical technique called renormalization group theory. It is similar in form to the standard k- ϵ model, but the RNG k- ϵ model more accurate and reliable for a wider class of flows.

3.3 COMBUSTION MODEL

In the simulation, the EDC (eddy-dissipation-concept) model was used to be the combustion model. The EDC model is an extension of the eddy-dissipation model to include detailed chemical mechanisms in turbulent flows. It assumes that reaction occurs in small turbulent structures, called the fine scales. The length fraction of the fine scales is modelled as Equation (5):

$$\xi^* = C \xi \left| \frac{v \epsilon}{K^2} \right|^{1/4}, \tag{5}$$

where * denotes fine-scale quantities, ζ is micro-scale number, C_ζ is a volume fraction constant equal to 2.1377, ν is kinematic viscosity.

Species are assumed to react in the fine structures over a time scale as equation (6) [15, 16]

$$\tau^* = C_\tau \left(\frac{\nu}{\varepsilon} \right)^{1/2}, \tag{6}$$

where C_τ is a time scale constant equal to 0.4082.

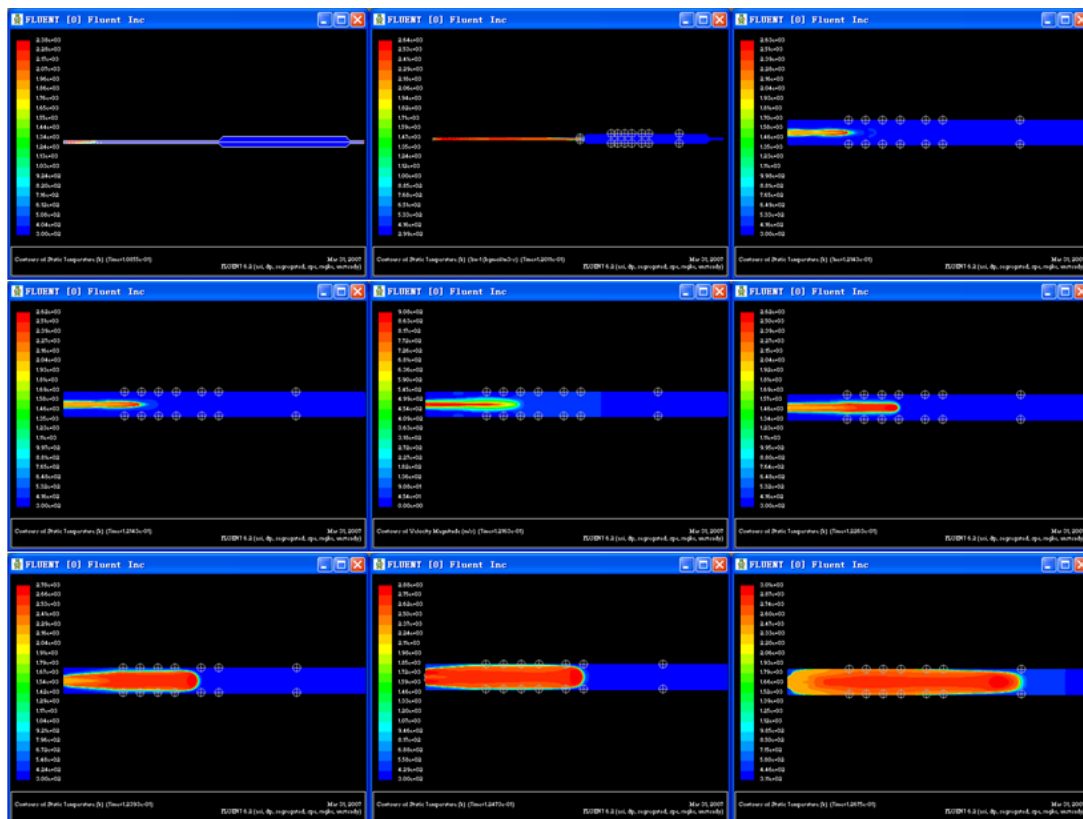
4 Boundary conditions

The duct wall-surface is considered as the outer boundary, and the open end of the duct is considered as the outflow boundary. There is no quality through the duct wall-surface. The whole calculated field is filled with 9.5% premixed methane-air mixture. At the initial moment, the premixed mixture is quiescent, and the premixed mixture temperature and pressure are 300 K and 1.0135×10^5 Pa, respectively. The velocity inlet, pressure outlet, and the porous jump boundary conditions are adopted [17].

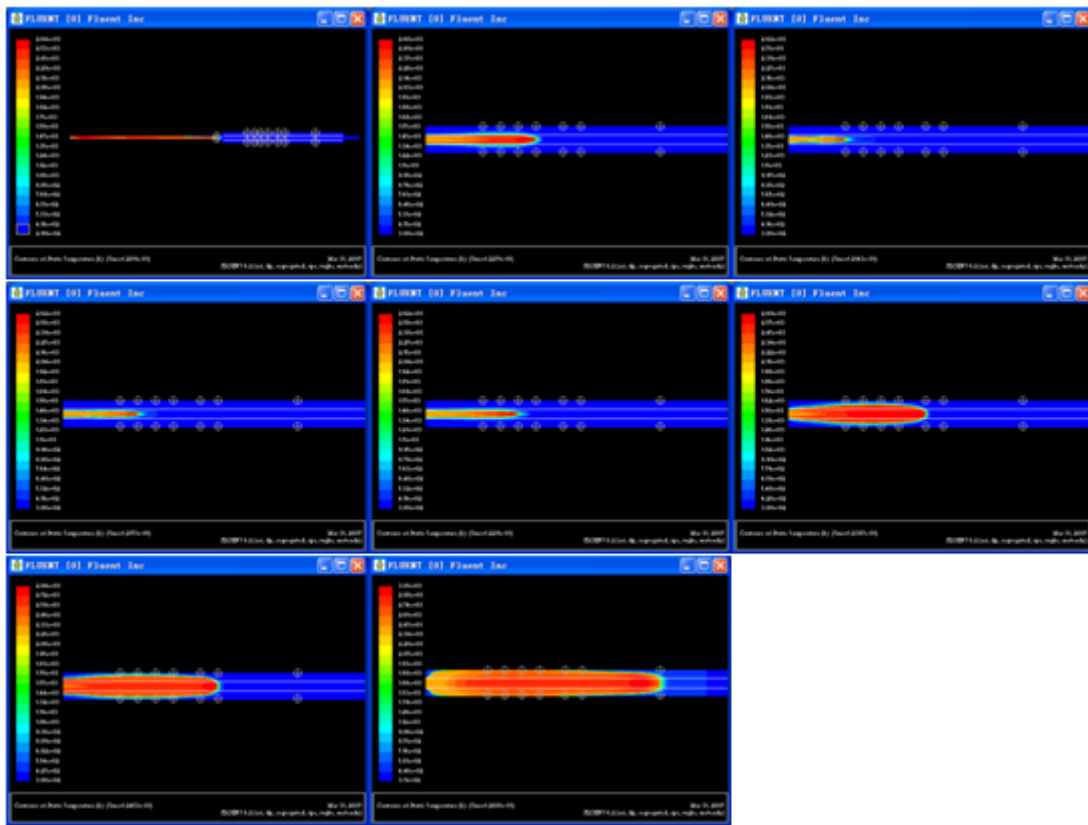
5 Results and Analysis on Numerical Simulation

5.1 NUMERICAL SIMULATION RESULTS

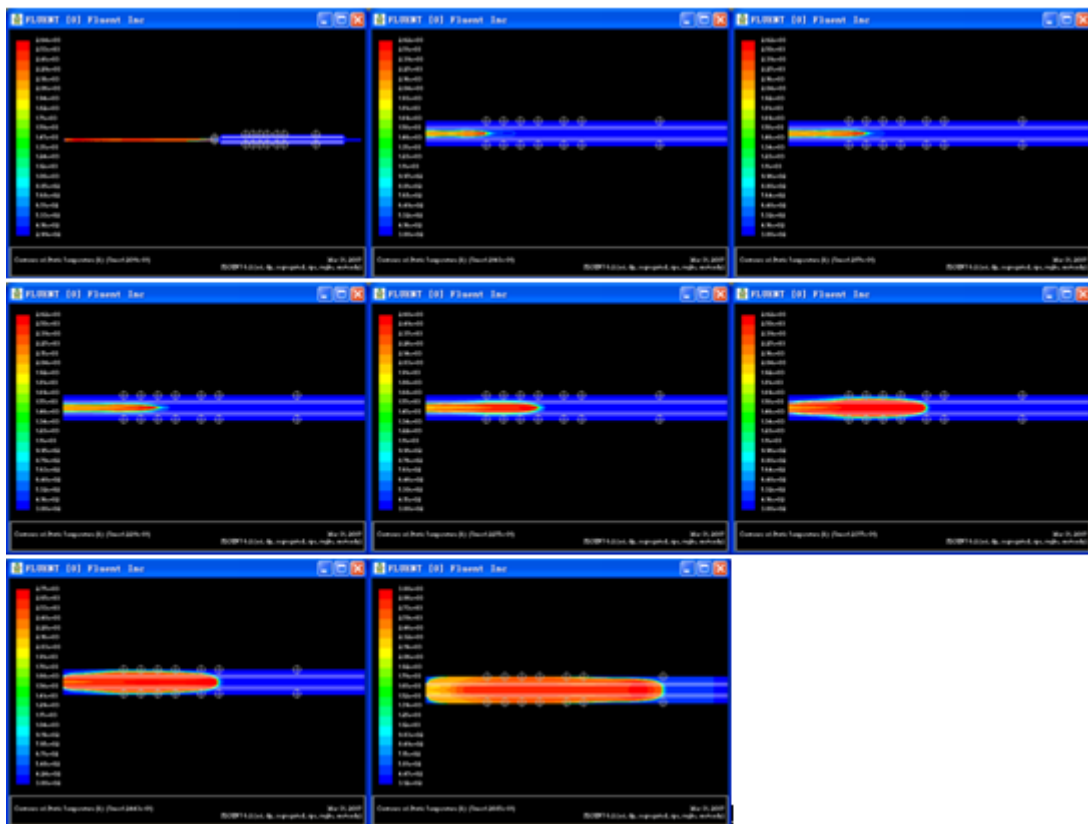
The theoretical and experimental results indicate that the methane concentration has great influence on the methane explosion intensity. When the methane concentration is 9.5%, methane explosion is most intense, and the flame propagation velocity, explosion shockwave pressure and temperature are biggest. Therefore, in the numerical simulation, the methane concentration in the premixed methane-air mixture was assumed to be 9.5%. The porous metal materials were installed in the range of 4.5-7.5 m away from the closed end of the duct, where the methane explosion intensity is greater. In order to facilitate the results analysis, eight pairs of measurement points are selected along the pipeline, the distances between them and the closed end of the duct are 4.3 m, 5.2 m, 5.4 m, 5.6 m, 5.8 m, 6.1m, 6.3 m, 7.2 m, respectively. In the process of numerical simulation, the explosion shockwave overpressure field, temperature field and flame motion field were calculated. The simulation results are shown in Figure 2 and Figure 3.



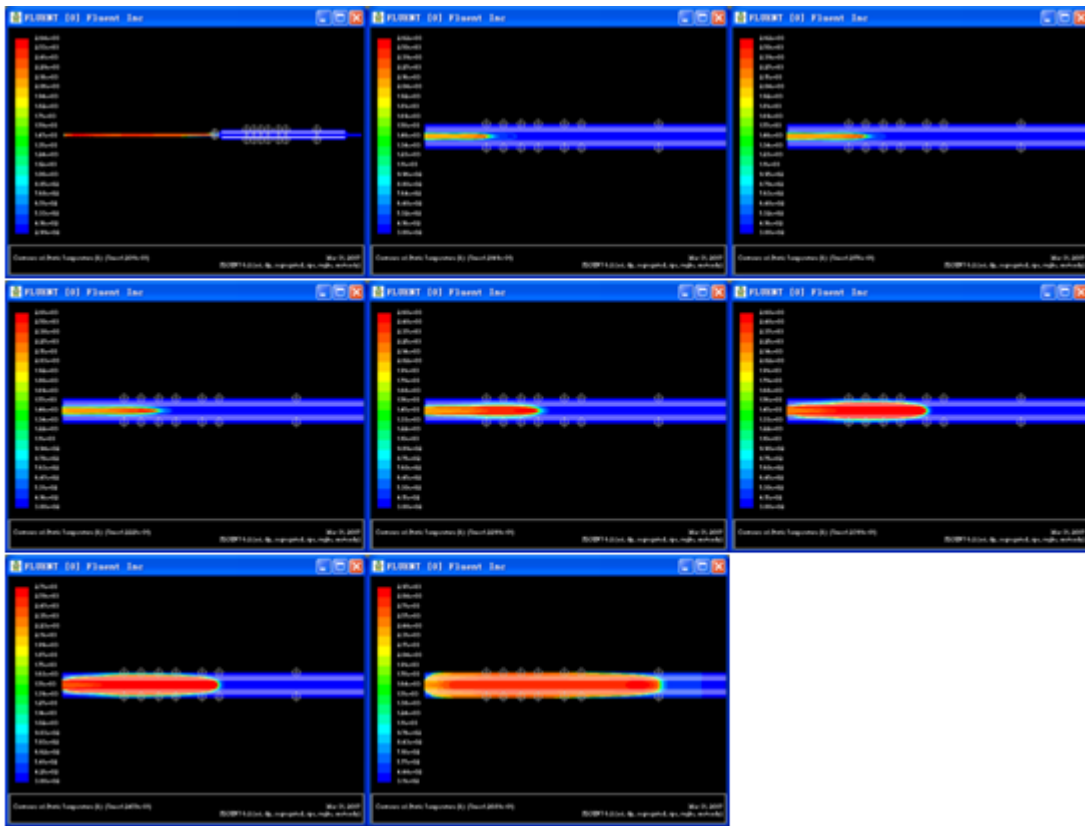
a. Duct without porous metal materials



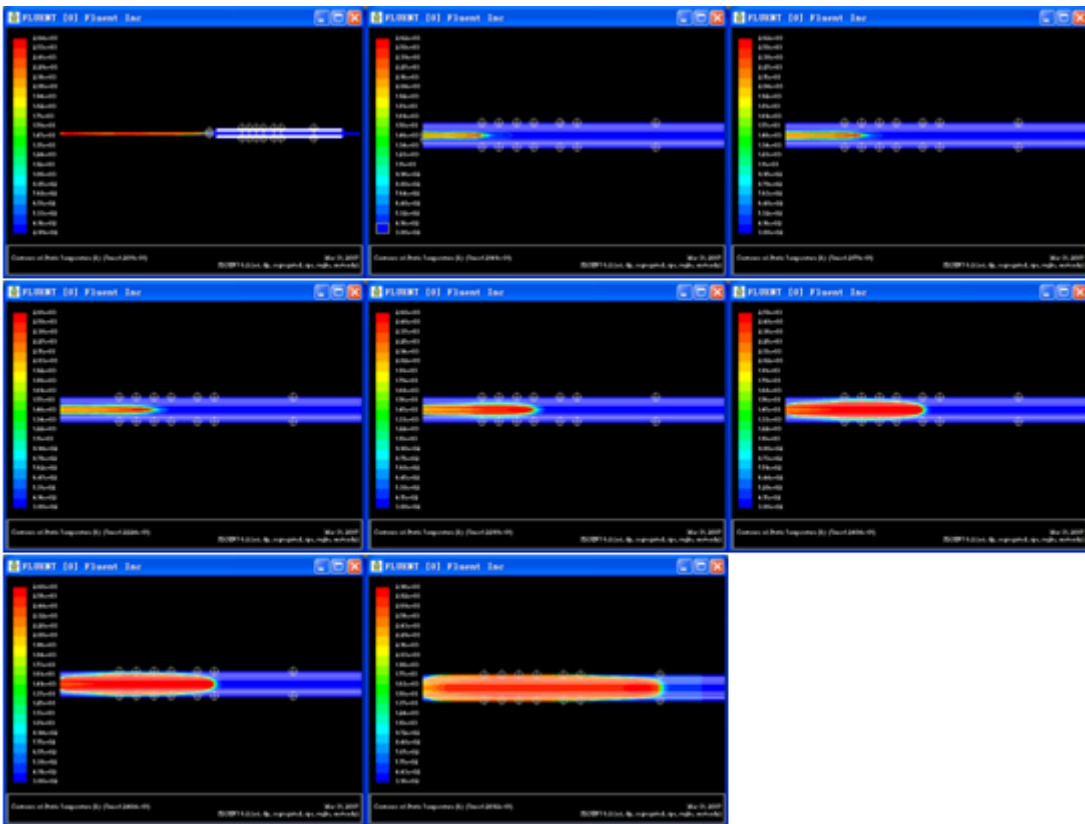
b. Duct with a layer of porous metal materials



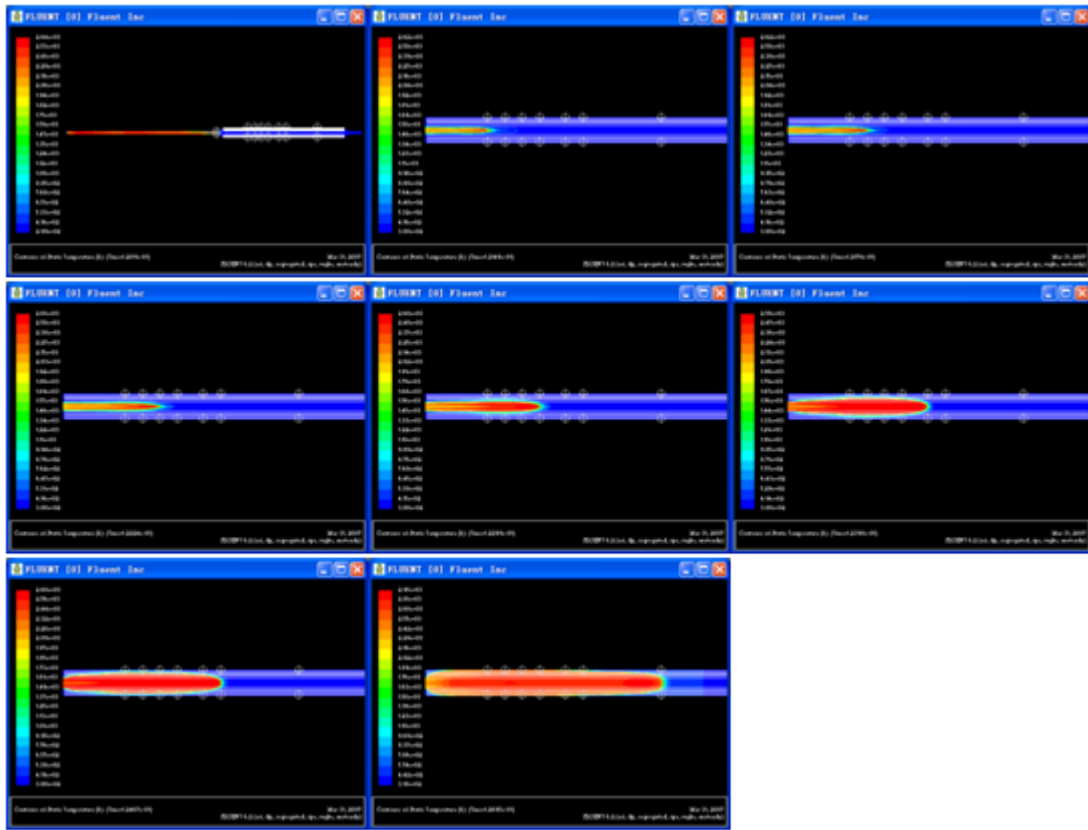
c. Duct with two layers of porous metal materials



d. Duct with three layers of porous metal materials



e. Duct with four layers of porous metal materials



f. Duct with five layers of porous metal materials

FIGURE 2 Flame propagation process of methane explosion in the duct with different layers of porous metal materials

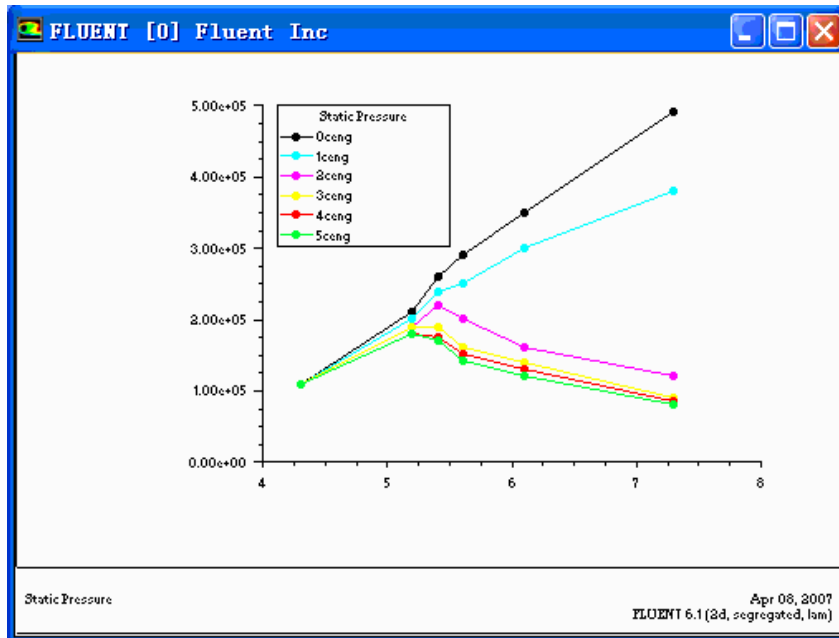


FIGURE 3 Overpressure-propagation distance scatter figure of methane explosion in duct with different layers of porous metal materials (the number of layers is equal to 0, 1, 2, 3, 4, 5, respectively)

5.2 ANALYSIS ON SIMULATION RESULTS

5.2.1 Analysis on flame propagation

From Figure 2, when methane explosion propagates in the duct with different layers of porous metal materials,

the flame propagation times (the flame propagation time refer to the time that the flame front at the axial reaches a certain place) from the ignition system to the same place of the duct are different. Regardless of the layers of porous metal materials, the flame propagation times from the ignition system to the second measurement point are

0.12143 s. Behind the second measurement point, the flame propagation time from the ignition system to the same measurement point increases with the increase of layers. When the duct is without porous metal materials, the flame propagation time from the ignition system to the eighth measurement point is 0.12675 s. When one layer of porous metal materials is installed with the duct, the flame propagation time from the ignition system to the eighth measurement point is 0.12681 s. When two layers of porous metal materials are installed with the duct, the flame propagation time from the ignition system to the eighth measurement point is 0.12685 s. When three layers of porous metal materials are installed with the duct, the flame propagation time from the ignition system to the eighth measurement point is 0.12689 s. When four layers of porous metal materials are installed with the duct, the flame propagation time from the ignition system to the eighth measurement point is 0.12692 s. When five layers of porous metal materials are installed with the duct, the flame propagation time from the ignition system to the eighth measurement point is 0.12695 s. Hence more layers of porous metal materials can bring lower flame propagation velocity.

Regardless of layers of porous metal materials, during flame propagation in porous metal materials, the flame velocity increases within a certain distance, which is considered as the minimum attenuation distance of the flame propagation velocity. Only the flame propagation distance is beyond the minimum attenuation distance, the flame propagation velocity begins to decrease. The numerical simulation results shows that the minimum attenuation distance is about 5-6 times the diameter of the duct, and it increases with the decrease of layers. Moreover, layers of porous metal materials have greatly effect on the flame propagation velocity, the flame propagation velocity decreases with the increase of layers. Therefore, the porous metal materials have good suppression effect on flame propagation in the process of methane explosion.

5.2.2 Analysis on explosion shockwave overpressure

When methane explosion propagates in porous metal materials, the repeated oscillations of explosion shockwave make the overpressure take on several peaks at the same measurement point, so that the *Fluent* software simulation figure is difficult to directly display the change process of explosion shockwave overpressure. Thus the post-processing is carried out by *Fluent* software, a XY scatter figure of overpressure-propagation distance of methane explosion in the duct with different layers of porous metal materials (the layers is equal to 0, 1, 2, 3, 4, 5, respectively) is drawn. As shown in Figure 3, methane explosion shockwave overpressure is affected by the layers and the length of porous metal materials. Only when the number of layers is more than 1, the explosion shockwave overpressure decreases obviously. The explosion shockwave overpressure is increasing within a

certain distance in porous metal materials, which is considered as the minimum attenuation distance of the explosion shockwave overpressure. Only the explosion shockwave propagation distance is beyond the minimum attenuation distance, the explosion shockwave overpressure decreases with the methane explosion propagation. The numerical simulation results shows that the minimum attenuation distance is about 5-6 times the diameter of the duct, and it decreases with the increase of layers. Moreover, layers have great influence on explosion shockwave overpressure. The more the layers are, the larger the decrease of the explosion shockwave overpressure is. Thus, the porous metal materials have good suppression effect on the explosion shockwave overpressure.

The numerical simulation results and their analysis show that, in the process of methane explosion propagation in a duct with porous metal materials, the attenuation of flame propagation velocity and explosion shockwave overpressure takes on synchronization and correspondence.

6 Conclusions

The methane explosion propagation in a one-dimensional straight duct with different layers of porous metal materials was numerically simulated, the conclusions are obtained as follows:

(1) The numerical simulation results can well reflect the process of methane explosion in porous metal materials, which show that the model is the correct and the numerical simulation is a good supplementary means of experiment.

(2) The porous metal material has good suppression effect on methane explosion propagation, the layers and the length of porous metal materials affect the flame propagation velocity and explosion shockwave overpressure. During methane explosion propagation in porous metallic materials, only the propagation distance is beyond the minimum attenuation distance, the flame propagation velocity and the explosion shockwave overpressure begin to decrease. The more the layers are, the better the attenuation effect is. Numerical simulation results are well consistent with the theoretical and experimental results. Hence porous metal materials can suppress methane explosion propagation.

(3) During methane explosion propagation in porous metal materials, the attenuation of explosion shockwave overpressure and flame propagation velocity takes on synchronization and correspondence.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (51004048, 51374003, 50834005, 50674047). Their supports are acknowledged with thanks.

References

- [1] Gaydon A G, Wolfhard H G 1970 *Flames: Their Structure, Radiation and Temperature (3rd ed.)* Chapman and Hall: London
- [2] Wang Z C, Terushige Ogawa 1995 The optimum choice of the design parameter on wire gauze *China Safety Science Journal* 5 (Supp.) 176-182 (in Chinese)
- [3] Bauer P 2005 Experimental investigation on flame and detonation quenching: applicability of static flame arresters *Journal of Loss Prevention in the Process Industries* 18(2) 63-8
- [4] Nie B S, He X Q, Zhang J F, Gu X M, Hu T Z 2008 The experiments and mechanism of foam ceramics affecting gas explosion process *Journal of China Coal Society* 33(8) 903-907 (in Chinese)
- [5] Gvozdeva L G, Faresov Iu M, Brossard J, Charpentier N 1986 Normal shock wave reflection on porous compressible materials *Progress in Astronautics and Aeronautics* 106 155-165
- [6] Dupre G, Peraldi O, Lee J H, Knystautas R 1988 Propagation of detonation waves in an acoustic absorbing walled tube *Progress in Astronautics and Aeronautics* 114 248-63
- [7] Vasil'ev A A 1994 Near limiting detonation in channels with porous Walls *Combustion, Explosion and Shock Waves* 30(1) 101-6
- [8] Teodorezyk A, Lee J H S 1995 Detonation attenuation by foams and wire meshes lining the walls *Shock Waves* 4(4) 225-36
- [9] Reddy K V, Fujiwara T, Lee J H 1988 Role of the transverse waves in a detonation wave – a study based on propagation in a porous wall chamber *Memoirs of the Faculty of Engineering, Nagoya University* 40(1) 149-62
- [10] Yu J L, Cai T, Li Y, Shi Q C 2008 Experiment to quench explosive gas with structure of wire mesh *Journal of Combustion Science and Technology* 14(2) 97-100 (in Chinese)
- [11] Fedorov A V, Fedorchenko I A 2010 Interaction of a normally incident shock wave with a porous material layer on a solid wall *Combustion, Explosion, and Shock Waves* 46(1) 89-95
- [12] Han F L 2008 *The Experiment and Numerical Simulation of The Flame Spread in Tubes with Porous Material* Dalian University of Technology: Dalian (in Chinese)
- [13] di Mare L, Mihalik T A, Continillo G, Lee J H S 2000 Experimental and numerical study of flammability limits of gaseous mixture in porous media *Experimental Thermal and Fluid Science* 21(1-3) 117-123
- [14] Zhao J H 2002 *Numerical Simulation of Combustion* Science Press: Beijing (in Chinese)
- [15] Hjertager L K, Hjertager B H, Solberg T 2002 CFD modeling of fast chemical reactions in turbulent liquid flows *Computers & Chemical Engineering* 26(4-5) 507-15
- [16] Gran I R, Magnussen B F 1996 A numerical study of a bluff-body stabilized diffusion flame Part 2 influence of combustion modeling and finite-rate chemistry *Combustion Science and Technology* 119 (1-6) 191-217
- [17] Liu R X, Wang Z F 2001 *Numerical Simulation and Dynamic Interface Tracking* Press of University of Science and Technology of China: Hefei (in Chinese)

Authors



Zhenzhen Jia, born in December, 1982, Xiangtan City, Hunan Province, P.R. China

Current position, grades: Lecturer of School of Energy and Safety Engineering, Hunan University of Science and Technology, China.

University studies: B.Sc. from China University of Mining and Technology in China, M.Sc. from China University of Mining and Technology in China.

Scientific interest: Gas Disasters Prevention in Coal Mines.

Publications: more than 10 papers published in various journals.

Experience: teaching experience of 7 years, 3 scientific research projects.



Feng Tao, born in April, 1957, Xiangtan City, Hunan Province, P.R. China

Current position, grades: Professor of School of Energy and Safety Engineering, Hunan University of Science and Technology, China.

University studies: B.Sc. from Chongqing University in China, M.Sc. from Chongqing University in China, D.Sc. from Central South University in China.

Scientific interest: Mining Engineering, Strata Control and Safety in Coal Mines.

Publications: more than 70 papers published in various journals.

Experience: teaching experience of 32 years, 20 about twenty scientific research projects.

Research and realization of handheld radio direction finding communication system

Bin Dong^{1*}, Guojie Yang¹, Tie Liang²

¹Hebei university Affiliated Hospital, Hebei University, Baoding 071002, China

²New Campus Management and Construction Office of Hebei University, China

Received 31 October 2013, www.tsi.lv

Abstract

Radio direction finding communication system has been widely used in civil and military fields. The handheld communication equipment with small development volume and low power has become the irresistible trend. Combined with magnitude comparison, a handheld direction finding and communication receiving system is researched and designed. And it is applied to radio communication system. This system can remove the influence of antenna on the hardware circuit, human body interference and surrounding environment. Besides, it has the advantages of small volume, fast direction finding speed and good capacity of resisting disturbance. Therefore, this handheld radio direction finding communication system has significant application value in public security, forest fire prevention and geological prospecting etc.

Keywords: radio, direction finding, communication, magnitude comparison

1 Introduction

There are various kinds of direction finding equipment selling in the market so far. For instance, the direction finding equipment for radio monitoring station adopts an antenna array on large basis. In addition, it processes the received signals from multichannel using high performance processor. It has the advantage of high measuring accuracy, but it also has obvious disadvantages, such as great costs and poor mobility etc. Currently, the disadvantages of some mobile measuring equipment are large volume and large power; therefore, it's difficult to cater to the requirements of handheld equipment. In addition, the hard operation is difficult for a user without actual operation experience. There is no communication function on the equipment. And it has no function of being found direction. Therefore, it cannot cater to search in the wild and gather the targets.

Based on the said problems, this article designs a handheld receiver with direction finding function after researching and concluding. This design can offset the influence on the antenna pattern caused by hardware circuit, human body interference and surrounding environment etc. Furthermore, it has the advantages of small volume, fast direction finding speed and good capacity of resisting disturbance. Therefore, it can be widely used for public security, fire prevention and geological resource exploration, etc.

2 Radio DF System

2.1 OVERVIEW OF RADIO DF SYSTEM

For the users, a suitable direction finding system is of great importance to their work. Therefore, how to choose a direction finding system match with their own work is the subject we are studying. When choosing the needed direction finding equipment, the user needs to consider the working conditions sufficiently, such as working condition, working mode, requirement, and object, etc. Instead, we cannot only talk about the advantages or disadvantages of the direction finding equipment. Only by choosing the suitable direction finding equipment for the working conditions, can its functions be given full play to. The direction finding systems in current market are as follows:

2.1.1 DF system of magnitude comparison

The working mode of direction finding system of magnitude comparison lies in that: when a direction finding equipment is proceeding, it will emit electromagnetic waves, and the receiver will judge the deviation angle by its direction finding antenna array or direction finding antenna. The amplitudes of electromagnetic waves received by the receivers from different directions are different. So the direction can be calculated by the equipment based on the magnitude of amplitudes. Actually, calculating direction by receiving electromagnetic waves is widely used in real life. There are also many direction-finding directional patterns of

* *Corresponding author* e-mail: dbin2000@163.com

different versions meanwhile. Loop, interval dual rings and rotating direction-finder antenna etc. among them just belong to directly rotating direction-finders. In addition, the commonly seen cross loop, U-shaped and H-shaped all belong to indirect rotating direction finders which realize the angle calculation by manual operating or circuit equipment.

There is one exception. The commonly seen Watson-Watt direction finder is a direction finding equipment of magnitude comparison, but it is not realized through magnitude comparison. Instead, it solves the direction angle of the incoming waves by calculating the arc tangent value.

2.1.2 Direction finding system of phase comparison

The working principle of this system is to calculate the direction angle according to the differences of voltage phases induced by the antenna, such as interferometer. It has obvious advantages of high sensitivity, high precision and fast speed etc. But its disadvantages are also obvious, that is poor capacity of resisting disturbance.

2.1.3 Direction finding system of difference in arrival time

The principle of this system is similar to the phase comparison, but the parameters they measure are different. Phase comparison measures time differences, while difference in arrival time measures phase differences. Direction finding system of difference in arrival time calculates the direction according to the time differences arriving at the equipment. This system needs hardware signal modulation.

2.1.4 Direction finding system of Doppler

According to the principle of Doppler Effect, when the equipment approaches to the emitter of electromagnetic waves at certain speed, the receiving frequency of the receiver will increase. When equipment leaves the emitter at certain speed, its receiving frequency will decrease. And when the equipment stops, its receiving frequency will stay unchanged. The direction finding system designed in accordance with such principle can fulfil the goal of direction finding. But it needs some small changes, the motion direction of the antenna shall be determined in line with the included angle between the circular motion direction of the equipment and the electromagnetic wave emitter. Its advantage is obvious with high precision and fast speed. However, it has poor capacity of resisting disturbance. Therefore, its application range is limited.

2.2 COMPARISON OF VARIOUS DIRECTION FINDING SYSTEMS

For users, it's of great importance to choose a suitable

system for their own work, so they need to comprehensively consider their own working condition and working requirement when choosing a direction finding system. It's not suitable for us to choose phase comparison system or Doppler direction finding system at fields where there is electromagnetic interference. But they can be adopted for fields with high requirement on precision and time. A kind of mobile and portable equipment is to be researched in the paper, which requires good motility of the system and excludes the possibility of using an antenna array on large basis (such as Wullenweber direction finding system). Considering from the respect of direction finding distance, this system requires remote direction finding and communication. So the direction finding system with bad direction finding distance such as the rotary antenna direction finding in magnitude comparison direction finding will not be used. Comparing from the respects of direction finding sensitivity and precision, the cross annular direction finding system in magnitude comparison direction finding system will not be chosen. Considering from capacity of resisting disturbance and frequency range, the U and H shaped antennas in fixed antenna array on small basis is excluded. Through the said comparison and based on our working conditions and requirement, Watson-Watt direction finding technology is the most appropriate choice since it is featured by broad frequency bandwidth, high sensitivity and high accuracy. Meanwhile, it is simple to make with small volume, light weight and sound motility.

2.3 PRINCIPLE OF MAGNITUDE COMPARISON DIRECTION FINDING TECHNOLOGY

2.3.1 Frequency scanning

Handheld equipment is used to scan the electromagnetic waves automatically. According to the software design, start and end frequency and step length are expressed by hexadecimal. There is no start and end bits for data. Therefore, in line with these two bits, data validation is conducted by the device [11]. Followed by the analysis of the received data, the signal strength is obtained to determine the frequency location of the handheld device. In the data analysis process, when the handheld device receives a concave projection, then the first received peak is the frequency of the device [12]. Based on the above analysis, a frequency scanning algorithms can be designed to determine the frequency of the handheld device.

2.3.2 Direction finding algorithm

Figure 1 is a direction finding schematic diagram of amplitude. In Figure 2, d , θ and λ represent the distance of two array elements, angle of incident signal and the carrier length in the signal receiving direction, respectively. Direction finding angle is $[-180^\circ, +180^\circ]$. As

seen from Equation (1), when $d < \lambda/2$, angle of incident signal θ is between $[-90^\circ, +90^\circ]$, and the phase difference is $[-180^\circ, +180^\circ]$. In a similar way, when $d > \lambda/2$ and the angle of incident signal θ is between $[-90^\circ, +90^\circ]$, the obtained phase difference may not be between $[-180^\circ, +180^\circ]$, and therefore, the phase difference may correspond to different values of incident angles. "Phase ambiguity" is called in the paper. Therefore, the value of d must be less than $\lambda/2$ so as to ensure the mapping between the actual orientation angle and measure angle.

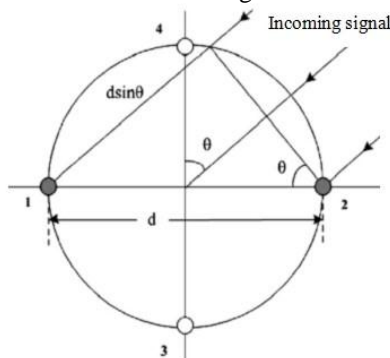


FIGURE 1 Measure principle of amplitude

Assuming that a plane electromagnetic wave is connected to an antenna in the angle of θ and with the incident distance of d , the phase difference can be obtained by formula (1):

$$\psi_{12} = \frac{2\pi d \sin \theta}{\lambda} \tag{1}$$

From the above equation, the corresponding angle of incidence θ can be calculated when we measure ψ_{12} .

Assuming the receive responsiveness of these two signal channels are the same, and the output phase difference of the receiving apparatus is φ , then the signal of the receiving apparatus is:

$$s_1(n) = U_1 \cos(n\omega + \psi_1), \tag{2}$$

$$s_2(n) = U_2 \cos(n\omega + \psi_2), \tag{3}$$

The receiving data model of receiving device will process obtained data of two antennas. Through the transformation of Hilbert, the sine-related expression can be obtained. And then the phase difference can be got by using the trigonometric function.

Supposing $x(n)$ gets $\hat{x}(n)$ after the transformation of Hilbert, then:

$$\hat{s}_1(n) = U_1 \cos(n\omega + \psi_1), \tag{4}$$

$$\hat{s}_2(n) = U_2 \cos(n\omega + \psi_2). \tag{5}$$

By virtue of Equation (4):

$$\sin(\psi_1 - \psi_2) = \sin \psi_1 \cos \psi_2 - \cos \psi_1 \sin \psi_2. \tag{6}$$

The phase and sine values of two antennas are calculated as:

$$\psi_{12} = \sin^{-1}(\hat{s}_1(n)s_2(n) - s_1(n)\hat{s}_2(n)). \tag{7}$$

Direction angle can be obtained by the formula (1). To solve the problem of phase ambiguity, data acquired from antenna 3 and 4 is analysed by the receiving apparatus. In this way, the obtained result can be more accurate. Then the phase difference and sine can be thus got.

3 System model design

The system is designed according to the magnitude comparison aforesaid. Currently, we have designed the system by means of the signals received by the antenna. All of these signals have chip MC3362 with indicator signal current RSSI:

$$RSSI = \varepsilon(10 \ln P_{in}). \tag{8}$$

In the formula above, ε represents the proportional constant of the chip. The following formula can be obtained based on Ohm's law:

$$P_{in} = \frac{V_{in}^2}{R_{in}}. \tag{9}$$

In the above formula, the input voltage is represented by V_{in} , and input impedance is represented by R_{in} , where the proportional relation of input voltage and the field strength at this point is as shown in the following formula:

$$V_{in} = KE. \tag{10}$$

In formula (10), K is the proportional constant. Substitute formula (9) and (10) into formula (4) to obtain the following:

$$RSSI = 20\varepsilon 10 \ln |E| + (20\varepsilon 10 \ln |K| - 10 \ln R_{in}). \tag{11}$$

For two orthogonal antenna devices, the error caused by the hardware devices can be neglected, through the difference of RSSI between the two channels, we can get:

$$|\tan(\theta)| = \frac{|E_2|}{|E_1|} = \exp\left(\frac{RSSI_2 - RSSI_1}{20\varepsilon}\right). \tag{12}$$

Seen from formula (11), the direction angle of the incoming wave is only related to the output current RSSI signal of the chips. The antenna direction finding device designed in this article can offset the error caused by hardware circuit and human factors. Thus the proportionality coefficient ε can be directly applied to the calculation. Then the direction angle of the wave can be calculated. That method has the advantages of high precision and good capacity of resisting disturbance.

Figure 2 shows the overall block diagram of the system designed by this article. This system consists of receiving module, loop receiving antenna, man-machine interface, microprocessor and phase-lock loop. The receiving module is used for receiving electromagnetic wave, and the man-machine interface is used for displaying the direction and strength of the signal, the microprocessor is used for processing the received signal and calculating the direction angle.

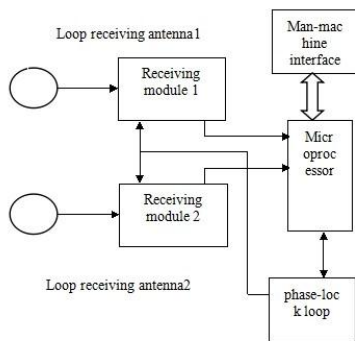


FIGURE 2 Block diagram of the hardware of radio direction finder

4 Experiment and analysis

4.1 TEST SYSTEM

As shown in Figure 2, the emitter in this system adopts FM modulation, with the bandwidth of 16KHZ, and power of 2W, the stability of the frequency not more than 20ppm, harmonic radiation not more than -40dBc. We input the induced signal into MC3362 and modulate the signal it receives, and then input it into the microprocessor for angle calculation after amplifying it in the operational amplifier, and finally display the operational amplifier and the calculated angle on the display screen.

4.2 TEST METHOD

We design a PC upper computer to analyse the data received, and establish a platform which can set the coils automatically. We send the data received via wireless module, so as to remove the interface of human factors, thus realizing the goal of precise data.

4.3 TEST DATA AND ANALYSIS

The actually tested directional diagram and the curve of amplitude comparison differences are shown in Figure 3 and Figure 4.

As shown in Figure 3, we can analyse the changing of the voltages received in the two channels. At the same moment, when the voltage in channel 1 is at the maximum point, the voltage in channel 2 will be at the minimum point, and vice versa. This shows the positional relation of the two antennas, i.e. they are orthogonal. At around 0° and 180°, channel 1 corresponds to the minimum point, which shows that the signal received at

this moment is the strongest. Meanwhile, we find that channel 2 corresponds to the maximum point at this time, which shows that the signal that channel 2 corresponds is the weakest. Also, we can find that at 90° and -75°, channel 1 and channel 2 correspond to the weakest and strongest signals respectively.

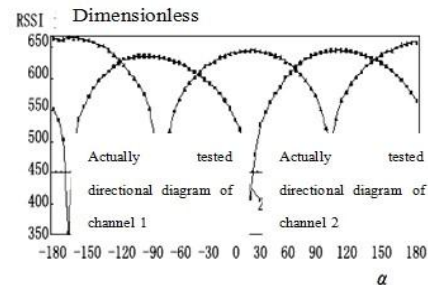


FIGURE 3 The actually tested directional diagram

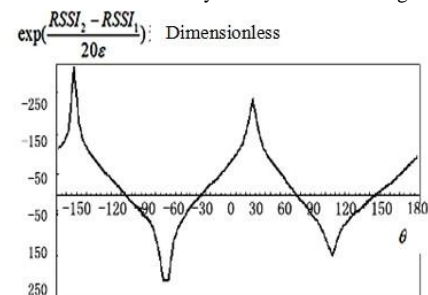


FIGURE 4 The actually tested curve of amplitude comparison differences

According to the analysis above, we can summarize the change rule of the indicator voltages, which basically complies with the pattern of orthogonal antenna. When the loop antenna 1 is at the minimum point, the change of the voltage field strength it corresponds to is very sensitive, therefore, we can completely set 0° and 180° as the basis to judge its direction. Meanwhile, we can also set it as the reference numerical value for distinguishing the direction of 0° and 180° according to the difference of the received signals.

Seen from the direction finding formula, we can find that its variable is the difference of the voltages indicated by the two channels, so we can conclude such conclusion: assuming the difference change significantly when changing in a certain direction, it will indicate that it has better sensitivity in this direction, therefore, this direction can be used for orientating. As shown in Figure 3, the magnitude of the difference between the voltages received by the two channels is shown. In the test, the difference values between the two channels change obviously when we rotate the direction finding antenna at somewhere near the maximum and minimum values. Therefore, it's easy for us to find the directions of the maximum and minimum values by changing the direction finding antenna, in particularly near 0°, so, we can find and determine the directions of the received electromagnetic wave manually. We can find the four directions with good sensitivity by further observation: they are the two maximum points and the two minimum

points respectively. Through Figure 4, we can also observe that there is a maximum point near 0° , through which we can find the direction of the wave by manual operation. In actual application, we need to calculate the directions continuously, therefore, the measurement accuracy can completely cater to the actual need of less than 0° .

TABLE 1 Table of relation between signal to noise ratio and angle error

Signal to Noise Ratio/dB	Angle error/degree	Signal to Noise Ratio/dB	Angle error /degree
0.0	5	4.0	3
0.5	5	5.0	3
1.0	5	6.0	2
1.5	5	7.0	2
2.0	4	8.0	2
2.5	4	9.0	1
3.0	3	10	1

To better reflect the direction finding accuracy of this design under different signal to noise ratios, in real application, the signal to noise ratio is measured by a noise factor analysis meter made by Agilent EXA at the reception end, so as to obtain the direction angle of the incoming wave by means of simple calculation in the PC through automatic testing system. The table of relation between signal to noise ratio and error of the wave of the

References

- [1] Jiang Xuejin, Gao Xia 2008 Design of Multi-Baseline Interferometer *Electronic Information Warfare Technology* **34**(4) 23-27
- [2] Chan Y T, Lee B H 2006 Direction Finding With a Four-Element Adcock Butler Matrix Antenna Array *IEEE Transactions on Aerospace & Electronics Systems* **37**(4) 1155-63
- [3] Zhai Qingwei, Wang Yu, Gong Bing 2007 Research on the Technology of the Beam Amplitude Comparison Direction Finding *Radio Communications Technology* **33**(6) 55-56
- [4] Holt Ollie Sampling of Comint/DF Receivers *Journal of Electronic Defense* 2007 **30**(7) 37-42
- [5] Li Yinbo, Chen Huajun 2008 Analysis and Application of Phase Detection Methods *Telecommunication Engineering* **66**(6) 34-38
- [6] Zhang Weihua, Zou Fang 2007 Influence of Difference of Amplitude and Phase between Channels to Interferometer Receiver and Its Solution Method *Journal of Electronics & Information Technology* **68**(5) 29-31
- [7] Benkert M, Gudmuordsson J 2008 *Constructing Minimum-Interference Networks Computational Geometry* **40**(3) 50-53
- [8] Jiang Xuejin, Gao Xia 2008 Design of Multi-Baseline Interferometer *Electronic Information Warfare Technology* **127**(4) 23-27
- [9] Herskovitz D 2005 A Sampling of Direction-Finding System *Journal of Electronic Defense* **23**(8) 57-66
- [10] Sun Hongxing, Fu Jianhong 2008 Ambiguity Resolution Based on Recursive Least Squares Kalman Filtering Using Multi-epoch Carrier Phase Data *Geomatics and Information Science of Wuhan University* **33**(7) 735-39
- [11] Miyamoto R 2007 A High-Direction Combined Self-Beam/Null-Steering Array for Secure Point-to-Point Communications *IEEE Transactions on Microwave Theory & Techniques* **55**(5) 838-44
- [12] Chen Zhan, Wang Xuesong, Tang Xuefei 2005 Short Wave Radio Finder System Based on Watson-Watt *Communications Technology* **35**(6) 25-27

Authors	
	<p>Bin Dong, born in July, 1980, Baoding City, Hebei Province, P.R. China</p> <p>Current position, grades: Advanced Engineer of HeBei University Affiliated Hospita, China. University studies: B.Sc. in Information and Computing Science from North China Electrical Power University of BaoDing in China. M.E. in Computer Science and Technology from North China Electrical Power University in China. Scientific interest: telecom, computer vision and medical information. Publications: about 10 papers published in various journals. Experience: engineering practice experience of 8 years in Hospital Information System, 6 scientific research projects.</p>
	<p>Guojie Yang, born in July, 1984, Baoding County, Hebei Province, P.R. China</p> <p>Current position, grades: Engineer of Hebei University Affiliated Hospital, China. University studies: B.Sc. in Electrical Engineering and Automation from Hebei University in China. M.Sc. from Hebei University in China. Scientific interest: information management, information fusion, virtual technology. Publications: more than 3 papers published in various journals. Experience: teaching experience of 3 years, 5 scientific research projects.</p>
	<p>Tie Liang, born in August, 1985, GongCheng County, GuangXi Province, P.R. China</p> <p>Current position, grades: Engineer of HeBei University, China. University studies: B.E. in Automation from HeBei University of BaoDing in China. M.E. from HeBei University in China. Scientific interest: wireless technology, automatic control. Publications: about 5 papers published in various journals. Experience: engineering practice experience of 3 years, 4 scientific research projects.</p>

Research on commutation torque ripple suppression strategy of BLDCM based on iterative learning

Jiong Zhang^{1*}, Shaofei Chen², Lihua Liu³

¹School of Vocational and Technical Education, Guangxi University of Science and Technology, 268 Dong Huan Road, Liuzhou Guangxi China

²Engineering Training Center, Guangxi University of Science and Technology, 268 Dong Huan Road, Liuzhou Guangxi China

³College of Science, Guangxi University of Science and Technology, 268 Dong Huan Road, Liuzhou Guangxi China

Received 4 April 2014, www.tsi.lv

Abstract

Brushless DC motor (BLDCM) is the DC motor which uses electronic commutation instead of mechanical commutation. The torque ripple caused by mechanical structure of the motor is small, while the torque ripple during the commutation is about 50% of the average torque. By analysing the unique features of commutation torque ripple, we can obtain the relationship between the commutation current and torque ripple. The current ripple can be suppressed through adding voltage compensation of the three-phase bridge inverter circuit. Then the commutation torque ripple suppression strategy based on iterative learning was raised in this paper. With the MATLAB Simulink platform, BLDCM simulation models, which based on the iterative learning are established and simulation contrast experiments of open or closed-loop iterative P-type with or without adding the voltage compensation have been designed to test and verify the effectiveness of the suppression strategy.

Keywords: BLDCM, Iterative learning, Torque Ripple, Voltage compensation

1 Introduction

The applied range of brushless DC motor (BLDCM) is becoming wider and wider because of its unique characteristics, such as small size, good performance, simple structure, high reliability, large output torque and so on. However, BLDCM cannot meet a number of occasions with high precision, for it has special problems of commutation torque ripple [1]. Therefore, the suppression of torque ripple is very important to the application of brushless DC motor in high-performance servo system.

Since the torque ripple of BLDCM is affected by electromagnetic factors, motor commutation, cogging effect, armature reaction and mechanical process factors, eliminating the torque ripple needs different methods in accordance with different causes. This paper only focuses on the main factor that causes torque ripple –the commutation. Corresponding voltage compensation is added at the moment of commutation based on characteristics of iterative learning algorithm to weaken the BLDCM commutation torque ripple [2].

2 Mathematical Model of Permanent-Magnet Brushless DC motor (PMBLDCM)

The spatial distribution of magnetic field of BLDCM's rotor is a square wave and the current waveform is also a square wave, while the back-EMF waveform is a 120° trapezoidal wave. In order to simplify the analysis, time-

domain equation of state is adopted [3]. And the following assumptions are also made to simplify the analysis:

- (1) The stator winding is a 60° full-pitch concentrated winding with star connection.
- (2) Eddy current and magnetic hysteresis losses are neglected and magnetic saturation is ignored.
- (3) The three-phase windings are completely symmetrical and armature reaction is not considered.
- (4) Self-inductance and mutual inductance of the stator winding are constants.
- (5) The top width of the air gap magnetic field is 120° electrical angle, the distribution is an approximate rectangular wave and armature reaction is ignored.
- (6) The switching characteristics of six power MOSFETS are ideal.

When the three-phase winding with star connection has no neutral line, it does not change with the position change of rotor for the reason of existence of rotor reluctance. Therefore, the coefficients of self-inductance and mutual inductance of stator winding are constants, the voltage balance equation of BLDCM [4-5] is:

$$\begin{bmatrix} u_A \\ u_B \\ u_C \end{bmatrix} = \begin{bmatrix} R_s & 0 & 0 \\ 0 & R_s & 0 \\ 0 & 0 & R_s \end{bmatrix} \begin{bmatrix} i_A \\ i_B \\ i_C \end{bmatrix} + \begin{bmatrix} L_s - L_m & 0 & 0 \\ 0 & L_s - L_m & 0 \\ 0 & 0 & L_s - L_m \end{bmatrix} p \begin{bmatrix} i_A \\ i_B \\ i_C \end{bmatrix} + \begin{bmatrix} e_A \\ e_B \\ e_C \end{bmatrix} \quad (1)$$

* Corresponding author e-mail:860454913@qq.com

where: u_a, u_b, u_c - winding voltage of each stator phase (V), i_a, i_b, i_c - winding current of each stator phase (A), e_a, e_b, e_c - winding EMF of each stator phase (V), R - resistance of each stator phase (Ω), L - self-inductance of each phase winding of the stator (H), M - mutual inductance between any two phase windings of the stator (H), P - differential operator. The electromagnetic torque equation is as below:

$$T_e = \frac{e_A i_A + e_B i_B + e_C i_C}{\omega / n_p} \tag{2}$$

where ω is the mechanical angular velocity of the rotor, T_e is the electromagnetic torque, n_p is the number of motor pole pairs.

3 Analysis of Ripple Torque

3.1 ANALYSIS OF THE COMMUTATION PROCESS

This paper only analyses non-overloaded commutation torque ripple problem [7-8], which meets the condition that $E > 0.25\alpha V$. Where E the back-EMF of the motor is, αV is the actual line voltage of the motor. It is assumed that the power supply voltage of the motor is constant during the entire running process. The equivalent circuit and bridge inverter circuit diagrams of BLDCM can be obtained by the formula (1) and they are shown in Figure 1-5.

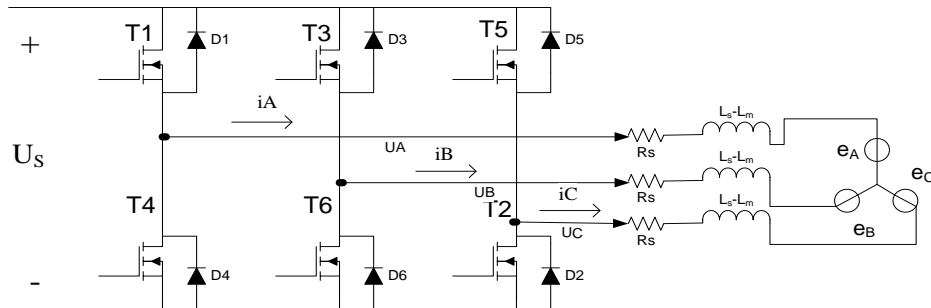


FIGURE 1 The equivalent circuit and bridge inverter circuit diagrams of BLDCM

where U_s is the supply voltage at DC side. T1-T6 are switching power Mosfets, D1-D6 are freewheeling diodes. U_A, U_B and U_C are phase voltages of phase A, B and C respectively. i_A, i_B and i_C phase currents of phase A, B and C respectively. R_s is interphase resistance. L_s-L_m is interphase inductance. e_A, e_B and e_C are electromotive forces of phase A, B and C. In this paper, we take the following case for example: the power supply of the motor changes from A-C phases to B-C phases.

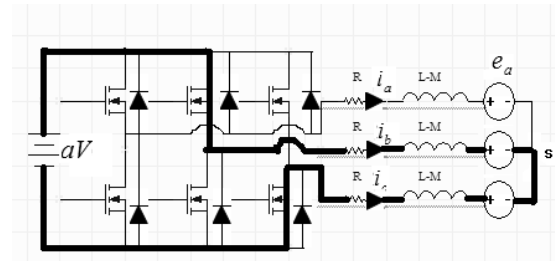


FIGURE 4 Mode after commutation

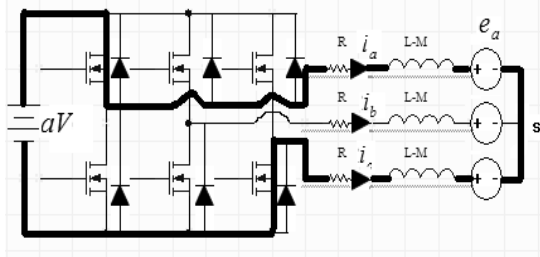


FIGURE 2 Single-current mode before commutation

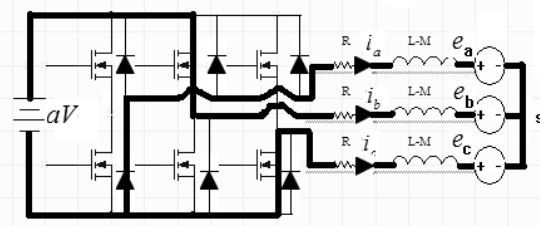


FIGURE 3 Current change mode during phase commutation

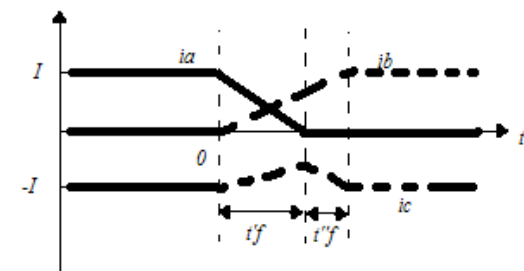


FIGURE 5 Current changes during the commutation process

3.2 CALCULATION OF COMMUTATION CURRENT

It can be found through the above analysis that the current commutation cannot be completed instantaneously in the commutation process due to the presence of inductance in armature winding. The current shutdown is delayed and needs the help of freewheeling diodes [9].

According to the Kirchoff's law, the influence of motor winding resistance on the commutation project is neglected because it is very small, and the circuit equation is:

$$\begin{cases} (L_s - L_m)pi_A + e_A - ((L_s - L_m)pi_C + e_C) = 0 \\ (L_s - L_m)pi_B + e_B - ((L_s - L_m)pi_C + e_C) = U_s \\ i_A + i_B + i_C = 0 \end{cases} \quad (3)$$

The waveform of back-EMF (back electromotive force) of each phase is a trapezoidal wave with top width of 120° electrical angle, i.e. $e_A = e_B = -e_C = E$. Considering that the initial and final values of each phase current are steady-state values before and after the commutation, the solution of Equation 3 is:

$$\begin{cases} i_A = I_p - \frac{U_s + 2E_p}{3(L_s - L_m)}t \\ i_B = \frac{2(U_s - E_p)}{3(L_s - L_m)}t \\ i_C = -I_p - \frac{U_s - 4E_p}{3(L_s - L_m)}t \end{cases}, \quad (4)$$

where I_p represents the steady-state value of the phase current.

3.3 RELATIONSHIP BETWEEN COMMUTATION CURRENT AND TORQUE

The three different situations in Figure 6 can be obtained based on the relations of the amplitude of i_A , i_B and i_C in formula (4).

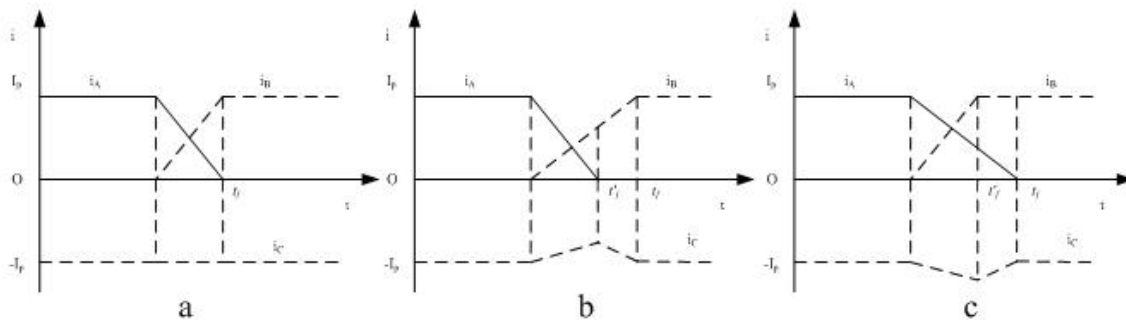


FIGURE 6 Relationship of the Commutation Current of BLDCM

(1) As shown in Figure 6 (a), i_B reaches steady state value, while i_A reduces to 0. Commutation time of A phase winding and B phase winding can be obtained by the formula (4), and they are equal. That is, the two phases will complete the commutation simultaneously, if they meet the following conditions:

$$U_s = 4E_p. \quad (5)$$

Since $i_A + i_B + i_C = 0$, the electromagnetic torque equation of commutation in process can be derived from formula (2) as below:

$$T_e = -\frac{2E_p}{\omega/n_p}i_c = \frac{2E_p}{\omega/n_p}(I_p + \frac{U_s - 4E_p}{3(L_s - L_m)}t) \quad (6)$$

Thus, the electromagnetic torque during commutation is proportional to the current of non-commutation phase winding. The electromagnetic torque during non-commutation period is generated by the interaction effect of synthetic MMFs of two phases and rotor permanent MMFs. It is shown as follows:

$$T_e = -\frac{2E_p}{\omega/n_p}I_p \quad (7)$$

By Equation (5), $U_s = 4E_p$, then:

$$T_e = T_e' - \frac{2E_p}{\omega/n_p}I_p. \quad (8)$$

So there is no torque ripple during the commutation process of BLDCM and this is an ideal work situation. The mechanical characteristic equation of BLDCM is:

$$n = \frac{U_s - 2\Delta U}{C_e\Phi_\delta} - \frac{2R_s}{C_e\Phi_\delta}I_i = \frac{U_s - 2\Delta U}{C_e\Phi_\delta} - \frac{2R_s}{C_T C_e\Phi_\delta^2}T_e, \quad (9)$$

where ΔU - Saturation voltage of power Mosfet, I_i - current of each phase winding, Φ_δ — magnetic flux per pole of corresponding square-wave air gap magnetic induction intensity, C_e - Electrical potential constant, C_T - Torque constant.

(2) As shown in Figure 6 (b), i_B has not reached its steady state value when i_A reduces to 0.

Similarly, it can be seen that the commutation of Figure 6(b) meets the following conditions:

$$U_s < 4E_p. \quad (10)$$

Formulas (6) and (7) show that at this moment. The electromagnetic torque in commutation process is lower

than the electromagnetic torque in non-commutation period. That is, commutation causes torque to decrease.

(3) As shown in Figure 6 (c), i_A has not reduced to 0 when i_B reaches its steady state value.

Similarly, it can be obtained that the commutation situation in Figure 6 (c) is opposite to that in Figure 6 (b). The electromagnetic torque in commutation process is bigger than the electromagnetic torque in non-commutation period. That is, commutation causes torque to increase.

4 Control Theory of Iterative Learning

4.1 MATHEMATICAL DESCRIPTION OF ITERATIVE LEARNING

Iterative learning control [10-11] (referred to as IILC) is a branch with strict mathematical description of intelligent control systems. It allows objects to track a desired trajectory with given accuracy within a given time zone, and it does not need identification of system parameters in the running process of the controller. Its mathematical description is as follows:

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) \\ y(t) = g(t, x(t), u(t)) \end{cases} \quad (11)$$

where $x(t) \in R^{m \times l}$ represent system state vectors, $y(t) \in R^{m \times l}$ represent the output vectors of the system,

$u(t) \in R^{m \times l}$ are the control vectors, f and g are vector functions of corresponding dimensions.

4.2 COMMUTATION TORQUE RIPPLE SUPPRESSION STRATEGY BASED ON ITERATIVE LEARNING

It can be obtained from section 2.3 that $U_s=4E$ must hold if no ripple is allowed during commutation. Positive voltage compensation is carried out when $U_s < 4E$ and negative voltage compensation is carried out when $U_s > 4E$. Torque ripple is actually current ripple. That is, during commutation, negative voltage compensation shall be employed when the detected current exceeds a certain value and positive voltage compensation shall be employed when the detected current is lower than a certain value so as to achieve the purpose of suppressing the torque ripple [12].

Iterative learning control algorithm is used as the speed controller. In addition, open and closed loop P-type iterative torque ripple suppression with voltage compensation and open and closed loop P-type control without voltage compensation are compared with each other to verify the effectiveness of this strategy. The diagram of open-loop P-type iterative torque ripple suppression with voltage compensation is shown in Figure 7. The diagram of closed-loop P-type iterative torque ripple suppression with voltage compensation is shown in Figure 8.

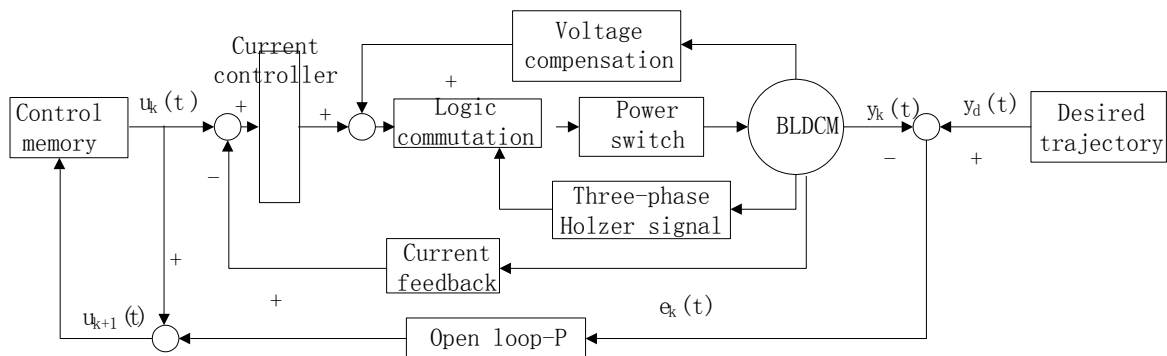


FIGURE 7 torque ripple suppression diagram of open-loop iterative P-type with voltage compensation

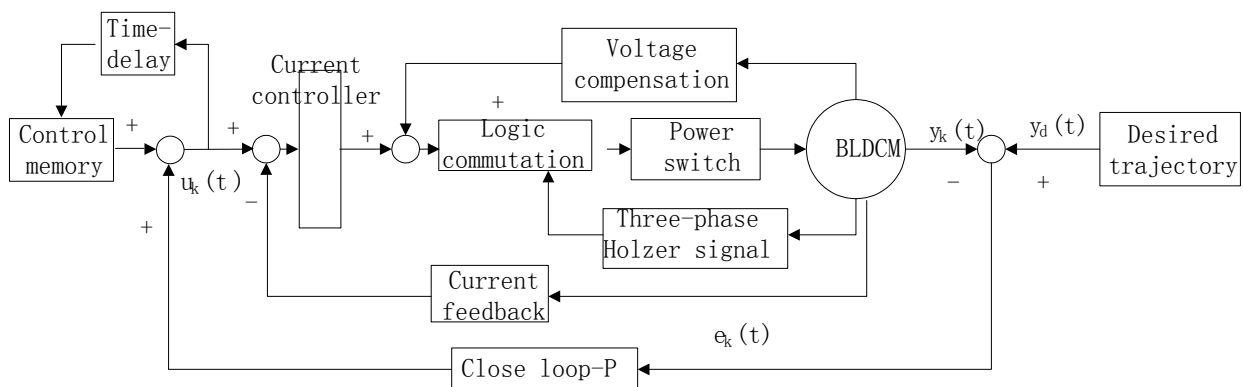


FIGURE 8 torque ripple suppression diagram of closed-loop iterative P-type with adding voltage compensation

5 Simulation of BLDCM Commutation Torque Ripple Suppression

5.1 SIMULATION MODEL OF DOUBLE CLOSED-LOOP CONTROL OF BLDCM

The simulation platform in this paper is MATLAB. Typical dual closed-loop control system of BLDCM is used and simulation parameters are set according to the data book of disc BLDCM provided by Maxon. Every independent sub-module is modelled separately according to the idea of modular modelling [13]. Four simulation models open and closed-loop P-type iteration without adding voltage compensation and open and closed-loop P-type iteration with adding voltage compensation are designed according to the commutation torque ripple suppression strategy based on iterative learning in section 3.

When the given velocity is 1000n/min, the given load is 0.03mNm and $t \in [0, 0.2]$, the initial conditions of the

system are $u_0(t) = y_d(t)$ and $u_0(t) = y_d(t)$,

$$y_d(t) = \begin{cases} 0 & t < 0.1 \\ 1000 & t \geq 0.1 \end{cases}$$

- (1) Open-loop iterative P-type control without adding voltage compensation
- (2) Closed-loop iterative P-type control without adding voltage compensation
- (3) Open-loop iterative P-type control with adding voltage compensation
- (4) Closed-loop iterative P-type control with adding voltage compensation

Open-loop P-type iterative simulation model without adding voltage compensation is shown in Figure 9, Close-loop P-type iterative simulation model without adding voltage compensation is shown in Figure 10, Open-loop P-type iterative simulation model with adding voltage compensation is shown in Figure 11, Close-loop P-type iterative simulation model with adding voltage compensation is shown in Figure 12.

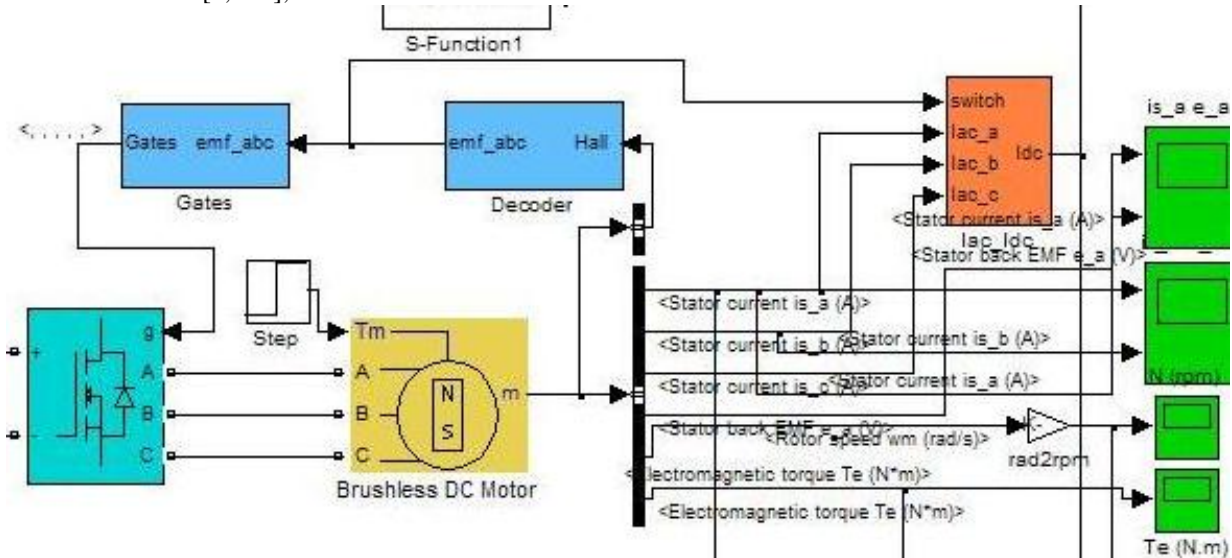


FIGURE 9 Simulation block diagram of open-loop iterative P-type control without adding voltage compensation

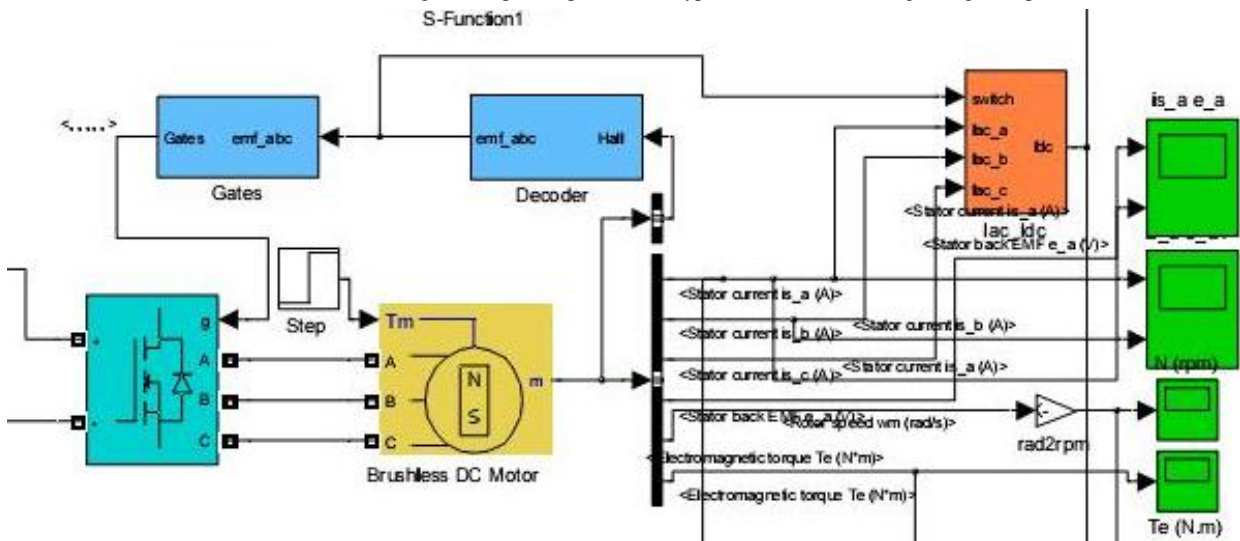


FIGURE 10 Simulation block diagram of the closed-loop iterative P-type control without adding voltage compensation

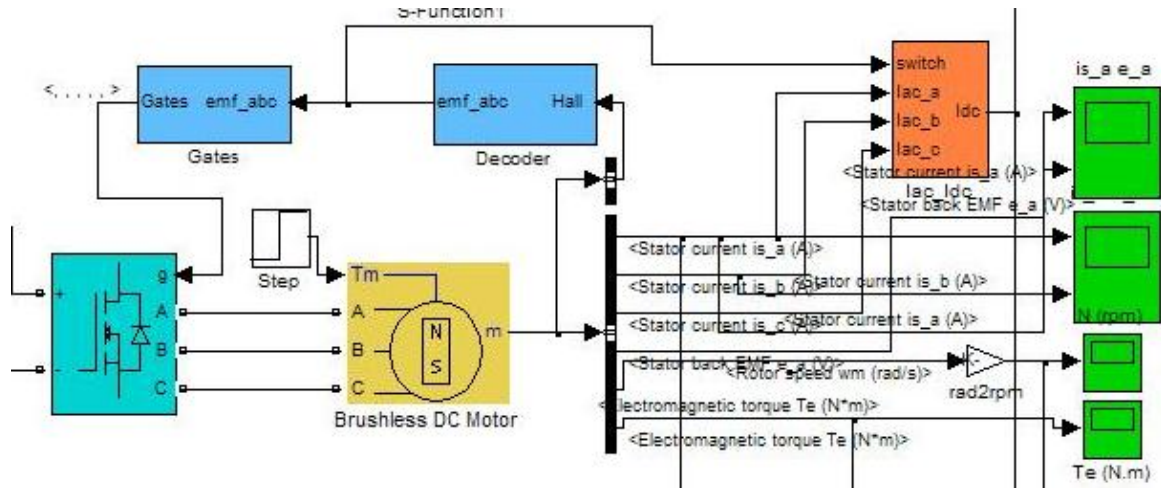


FIGURE 11 Simulation block diagram of the open-loop iterative P-type control with adding voltage compensation

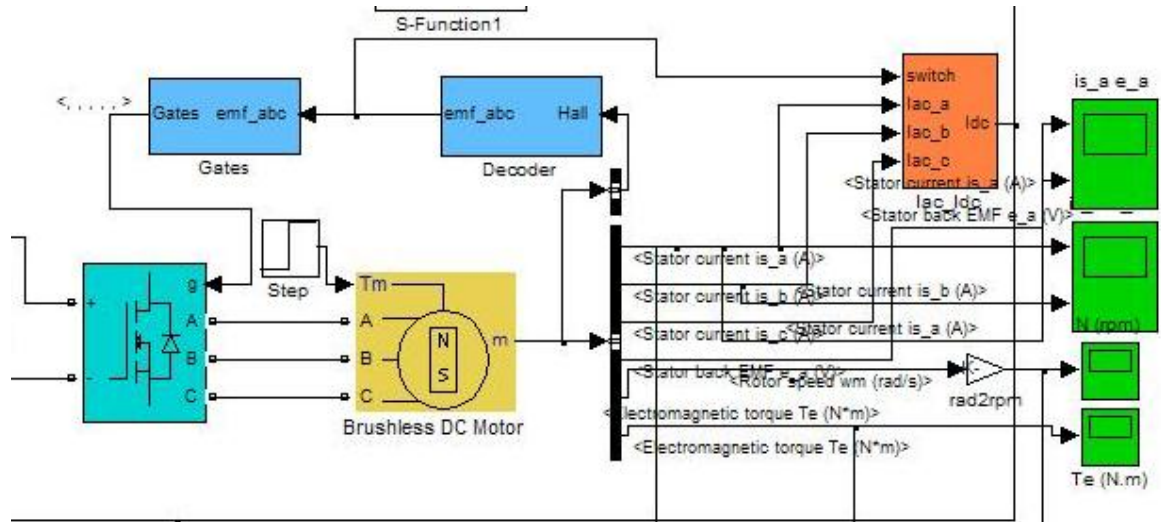


FIGURE 12 Simulation block diagram of the closed-loop iterative P-type control with adding voltage compensation

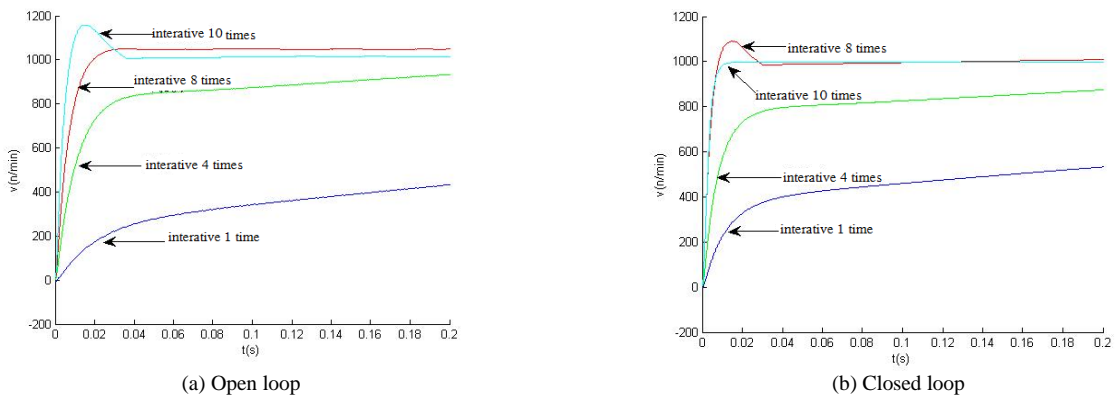


FIGURE 13 Speed waveform of open-loop iteration P-type control without adding voltage compensation

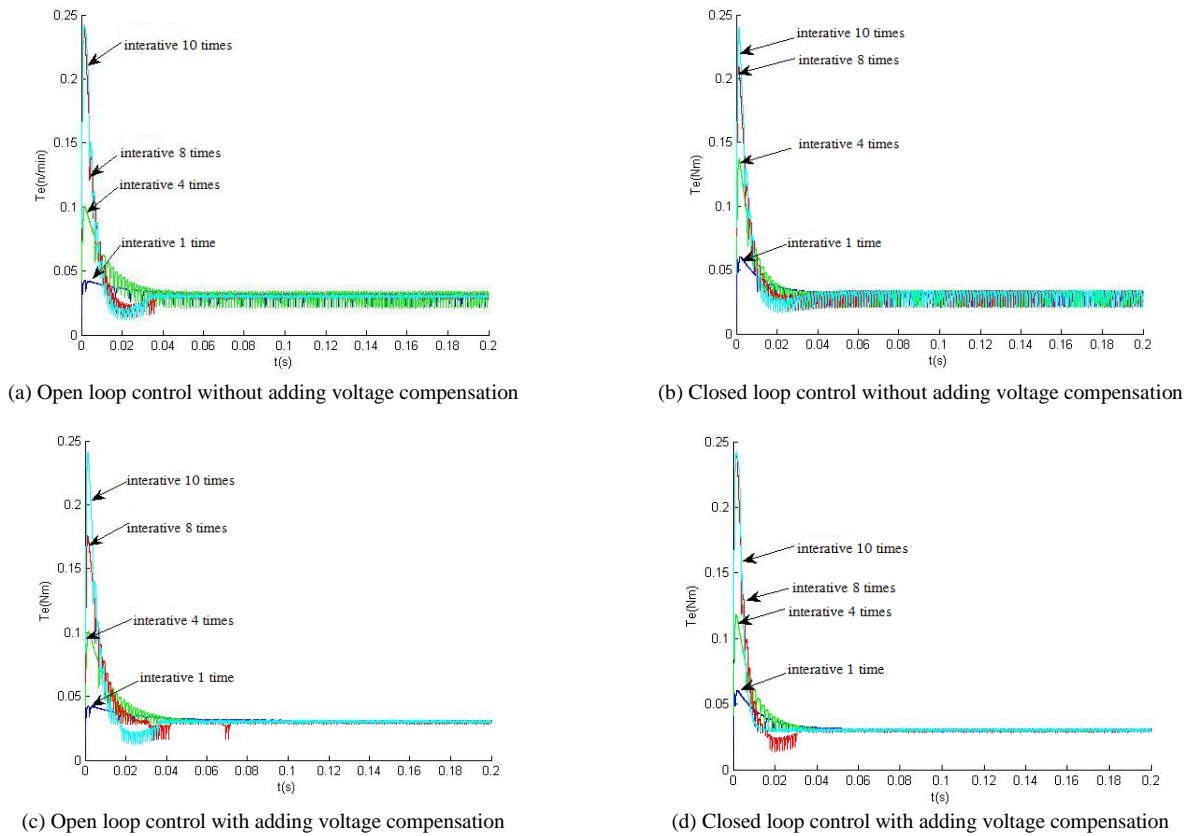


FIGURE 14 Torque ripple diagram of iterative P-type control

5.2 SIMULATION ANALYSIS

(1) Comparison between open-loop and closed-loop P-type iterative control method

From the comparison between (a) and (b) in the Figure 13, it can be seen that the step response of the control system under closed loop at 10th iterative learning step almost has no overshoot and is much faster than that of control system under open loop control.

TABLE 1 Parameters compare of two control methods

Control Method	Dynamic Performance		
	$\sigma\%$	t_r	t_s
open-loop P-type	18%	0.018	0.024
closed-loop P-type	0%	0.011	0.011

(2) In closed-loop P-type iterative control, comparison between torque ripple with adding voltage compensation control and torque ripple without adding voltage compensation control is shown in Table 2. By the comparison of four pictures in Figure 14, it can be seen that the torque fluctuation range without adding voltage compensation is [0.02, 0.033] and the torque fluctuation

range with adding voltage compensation is [0.029, 0.0312]. The torque ripple drops to 4% from 33.3%. The torque ripple partial enlarged drawing of closed loop iteration P-typed control without adding voltage compensation is shown in Figure 15. The torque ripple partial enlarged drawing of closed loop iteration P-typed control with adding voltage compensation is shown in Figure 16.

TABLE 2 Comparison of with and without adding voltage compensation

Control Method	Dynamic Performance	
	torque fluctuation range	torque fluctuation percentage
With voltage compensation	[0.02, 0.033]	33.3%
with voltage compensation	[0.029, 0.0312]	4%

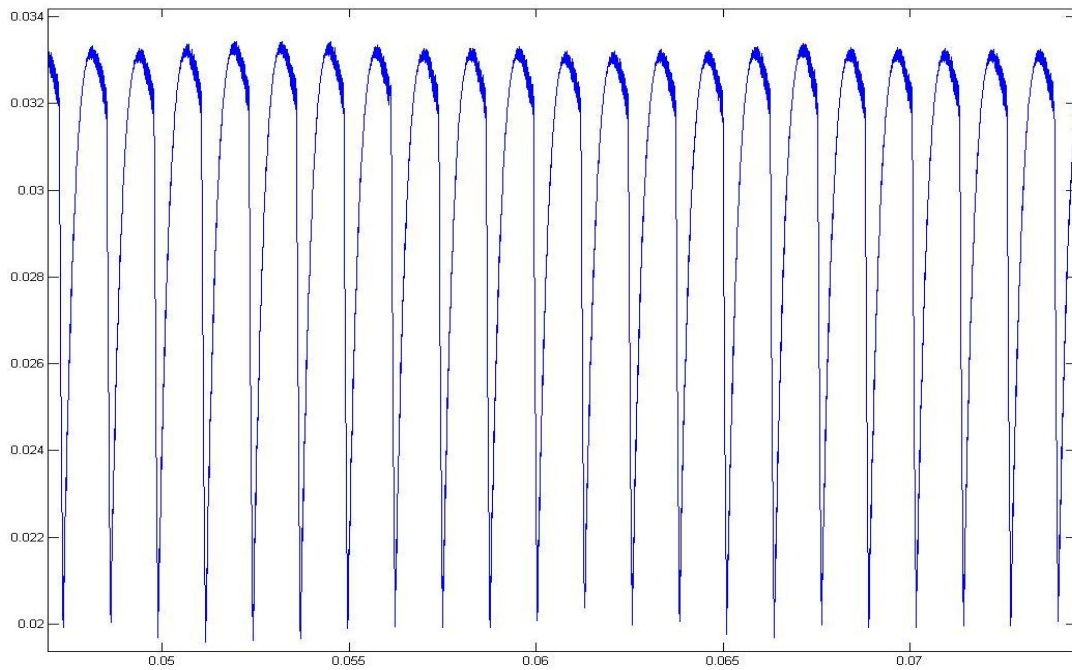


FIGURE 15 Torque ripple partial enlarged diagram of closed-loop iterative P-type control without adding voltage compensation

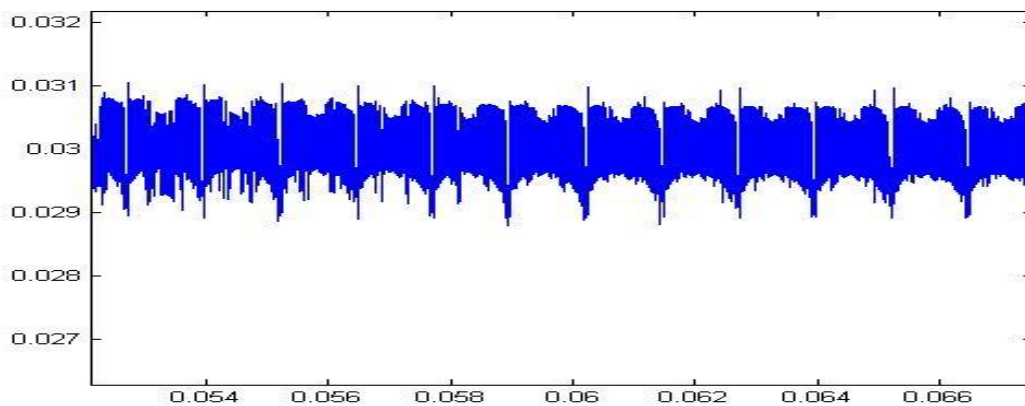


FIGURE 16 Torque ripple partial enlarged diagram of closed-loop iterative P-type control with adding voltage compensation

By analysing the above simulation experiments, it is obtained that the speed control performance of BLDCM under closed-loop iterative P-type control is better than that under open-loop P-type iterative control. With adding voltage compensation to the commutation torque ripple, the torque ripple drops to 4% from the original 33.3%. It means that the torque ripple with adding voltage compensation is greatly restrained. It verifies the effectiveness of commutation torque ripple suppression strategy of BLDCM based on iterative learning control [14-15].

6 Conclusions

Commutation torque ripple is the main factor that restricts the applications of permanent magnet brushless DC motor in high performance alternating current

governor systems. Therefore, this paper takes commutation torque ripple of BLDCM as the research object, obtains the relationship between the commutation current and the torque ripple through the analysis of the commutation process and the calculation of commutation current and raises the commutation torque ripple suppression strategy based on the iterative learning. This strategy is realized by adding voltage compensation during commutation of BLDCM based on the detected conditions of current ripple. Through the design of four simulation models and analysis of simulation experimental data, it is obtained that the torque ripple reduces to 1/8 of its previous level by adding voltage compensation during commutation and the effectiveness of commutation torque ripple suppression strategy of BLDCM based on iterative learning is therefore verified.

References

- [1] Zwahlen R, Chang T 2007 Feed forward Speed Control of Brushless DC Motors with Input Shaping The 33rd Annual Conference of the *IEEE Industrial Electronics Society(IECON)* 1169-74
- [2] Zhao Yao, Wang Huizhen, Chen Kang 2013 A Study About the Control System of BLDCM *Transactions of China Electrotechnical Society* **28**(1) 321-5
- [3] Liu Dongliang, Cui Yanfei, Chen Meibin 2013 Research of Back-EMF Estimation for Brushless DC Motor *Transactions of China Electrotechnical Society* **28**(6) 52-8
- [4] Peng Bing, Xia Jiakuan, Wang Chengyuan, Jing Rubao 2012 Analysis and Calculation of Zigzag Leakage Flux in Surface-Mounted PM Synchronous Machines with Similar Number of Poles and Slots *Transactions of China Electrotechnical Society* **27**(1) 114-8
- [5] He Hucheng, Liu Weiguo, Lang Baohua 2007 Research on Torque Characteristic of Non-commutation in PM Brushless DC Motor. *Electric Drive* **37**(10) 114-8
- [6] Pandit M, Buchheit K H 1999 Optimizing Iterative Learning Control of Cyclic Production Process with Application to Extruders *IEEE Trans* **7**(3) 382-90
- [7] Zhu Chi, Aiyama Yasumichi, Arai Tamio, Kawamura Atsuo 2006 Positioning in Releasing Manipulation by Iterative learning Control *Journal of Intelligent and Robotic Systems* **46**(4) 383-404
- [8] Sun Mingxuan, Wang Danwei 2000 Iterative learning Control Design for Uncertain Dynamic Systems with Delayed States *Dynamics and Control* **10**(4) 341-57
- [9] Ionel D M, Eastham J F, Betaer T 1995 Finite element analysis of a novel brushless DC motor with flux barriers, *Magnetics IEEE Transactions on* **31**(6) 3749-51
- [10] Jeon Y S, Mok H S, Choe G H, Kim D K, Ryu J S 2000 A new simulation model of BLDC motor with real back EMF waveform, *Computers in Power Electronics the 7th Workshop on* 16-18 July 2000 217-20
- [11] Zhu Z Q, Howe D 2000 Influence of design Parameters on toggging torque in permanent magnet machines *IEEE Trans on Energy Conversion* **43**(5) 407-11
- [12] Dai Min, Keyhani Ali, Sebastian Tomy 2001 Torque ripple analysis of a permanent Magnet brushless DC motor using finite element method *Proceedings from IEMDC* **32**(6) 241-5
- [13] Hwang S M, Lieu D K 1995 Reduention of torque ripple in brushless DC motors *IEEE Trans on Magneties* **31**(6) 3737-39
- [14] Kim Yoon-Ho, Kook Yoon-Sang, Ko Yo 1997 A new technique of reducing torque ripples for BLDCM drives *IEEE Trans.on Industrial Electronics* **44**(5) 735-9
- [15] Kim Yoon-Ho, Cho Byung-Guk, Ko Yo 1994 Generalized techniques of reducing torque ripples in BLDCM drives *Proceedings from IECON* **21**(2) 514-9

Authors

	<p>Zhang Jiong, born in February, 1978, Liuzhou County, Guangxi Province, P.R. China</p> <p>Current position, grades: lecturer of School of Vocational and Technical Education, Guangxi University of Science and Technology, China. University studies: B.Sc. in College of mechanical and power engineering from Chongqing University of Science and Technology in China. M.Sc. from Chongqing University in China. Scientific interest: Manufacturing Engineering and Automation, Materials processing engineering. Publications: more than 10 papers published in various journals. Experience: teaching experience of 7 years, 5 scientific research projects.</p>
	<p>Chen Shaofei, born in March, 1981, Liuzhou County, Guangxi Province, P.R. China</p> <p>Current position, grades: lecturer of School of Engineering Training Center, Guangxi University of Science and Technology, China. University studies: B.Sc. in Electrical Engineering and Automation from Hefei University of technology in China. M.Sc. from Chongqing University in China. Scientific interest: Control theory and control engineering, Electrical and electronic science. Publications: more than 4 papers published in various journals. Experience: teaching experience of 4 years, 3 scientific research projects.</p>
	<p>Liu Lihua, born in March, 1981, Liuzhou County, Guangxi Province, P.R. China</p> <p>Current position, grades: lecturer of School of College of Science, Guangxi University of Science and Technology, China. University studies: B.Sc. in College of Science from Guangxi University of Science and Technology in China. M.Sc. from Gullin university of electronic in China. Scientific interest: Management and optimization control, Management science. Publications: more than 5 papers published in various journals. Experience: teaching experience of 7 years, 4 scientific research projects.</p>

The gradual learning static load modelling method based on real-time fault recorder data

Guoping Shi^{1, 2*}, Jun Liang¹

¹*School of Electrical Engineering, Shandong University, Jinan City, China, 250061*

²*School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan City, China, 250101*

Received 1 April 2014, www.tsi.lv

Abstract

Setting a real-time load model is an effective way to overcome time-variation of power load in course of power load modelling. On the basis of load data sorting, this paper proposes a gradual learning static load modelling method based on power fault recorder data. Firstly, power fault recorder collects and stores valid load data. Secondly, all valid load data will be classified by the time, static load model can be built corresponds to each classification. Thirdly, model parameters of each sort are identified by gradual learning method, for the goal of global fitting optimal for the measured active power and calculated active power, the load model parameters are optimized by using curve fitting method. The identified model parameters can be applied to power system calculation directly without preserving all load data, essential feature of all load data is reserved and modelling operational efficiency is improved greatly. Simulation results show that the gradual learning method is right and effective, which is easier to realize and is of higher precision compared with least squares method, therefore the method has widely applicable value and is prospective in power system on-line static load modelling.

Keywords: Fault Recorder, Static Load Modelling, Parameter Identify, Gradual Learning, Curve Fitting

1 Introduction

The static load model is commonly employed in power system state analysis, such as power flow computing and some steady analysis based on power flow [1]. Large amounts of simulation results and tests showed: the static load model has great influence on power flow computing, voltage stability computing, frequency stability computing, reactive power compensation equipment planning, long time dynamic process analysing, and so on [2]. Under critical conditions some qualitative conclusions can even be reversed [3]. Therefore, it is highly necessary to research on static load modelling.

At present, the solutions to static load modelling in power system mainly include: statistical synthesis method, steady-state testing method and measurement-based method [4]. In recent years, appearance of various kinds of new electronic devices makes load characteristics become more and more complex, as time goes on, time-variant features of load are increasingly obvious [5]. According to field measurements made by American GE Company [6], static characteristic coefficient of an area changed 20% in 10 minutes. Static load modelling method based on one survey or one test can not reflect time-variant features of power load perfectly [7]. Therefore, to solve the problem above, the best way is on-line and real-time data acquisition, on-line data processing and on-line load model parameters identification. However, until now, on-line or real-time static load modelling method has not being used in power

system due to limitation of field measurement equipment.

With the rapid development and prevalence of power fault recorder, recorder makes on-line static load modelling possible [8]. This paper proposes a gradual learning static load modelling method based on real-time fault recorder data, together with the idea of statistical synthesis method. Power load field measured data is classified by their time characteristics, then load model parameters database can be gotten for building static load model corresponded with each class, according to real-time steady-state load data, different time-scale load models can be gotten through on-line correcting load model parameters database. Users choose these models according to their actual need; the time-variant characteristic of power load has been overcome. This paper builds and analyses a load model using measured summer load data of one day in Rizhao, experimental results verify the correctness and effectiveness of the method.

2 On-line static load modelling

Static load model is function equation between load power and voltage, frequency in steady-state conditions. The common static load models are polynomial model and power function model neglecting effects of frequency [9], shown in formula (1) and (2).

* *Corresponding author* e-mail: shiguoping@sdjzu.edu.cn

$$\begin{cases} P = P_0 \left[A_p (U/U_0)^2 + B_p (U/U_0) + C_p \right] \left(1 + \left(\frac{\partial P}{\partial f} \right)_{f_0} \Delta f \right) \\ Q = Q_0 \left[A_q (U/U_0)^2 + B_q (U/U_0) + C_q \right] \left(1 + \left(\frac{\partial Q}{\partial f} \right)_{f_0} \Delta f \right) \end{cases}, \quad (1)$$

$$\begin{cases} P = P_0 \left(\frac{U}{U_0} \right)^{p_u} \left(\frac{f}{f_0} \right)^{p_f} \\ Q = Q_0 \left(\frac{U}{U_0} \right)^{q_u} \left(\frac{f}{f_0} \right)^{q_f} \end{cases}. \quad (2)$$

Normally, voltage and power fluctuations are keeping small, the static load models are obtained by statistical synthesis method or steady-state testing method, nevertheless, these two methods mentioned above maybe cannot take place often, the static load model based on which cannot accurately describe the time-variant of load [10]. With the development and prevalence of fault recorder, it is possible for on-line real time static load modelling, on-line measuring load inputs and outputs data and building static load model can reflect time-variant feature of load and agree with the actual load [11].

2.1 LOAD DATA PROCESSING

Effective information load data through extracting from magnanimity data gathered by fault recorder can be used to build static load model. New-type fault recorder can realize the function of steady-state, transient-state, Long-term continuous dynamic process recording and analysing, can analyse and process the data acquired, and set judgment criterion, then qualitative data can be acquired for static load modelling conveniently.

2.2 CHOOSING LOAD MODEL STRUCTURE

There are polynomial model, power-function model and permutations of these two models [12]. With increasingly complex load characteristic, nonlinear functions load model structure has arisen, such as spline functions model, neural network etc [13]. Polynomial model structure is carried out in this paper.

2.3 MODEL PARAMETERS IDENTIFICATION ALGORITHMS

Model parameters need to be identified by measured recorder data after load model structure was selected. Fault recorder gathered load data continuously, if each time load data new acquired combined with old load data are used in identifying model parameters, the identifying process will be a long and hard one, the load data got by recorder need refresh constantly, moreover, recorder

storing capacity is limit, it is not feasible to identify static load model parameters making use of all the load data. This paper proposes an on-line recurrence revision parameters identifying method: only store load model parameters identified results, when a new set of load data is collected, the results will be revised on basis of original results. The approach need less calculation, compute with high efficiency, and extract essential feature of all load data, through which compose static load model describing all load data samples can be get.

2.4 LOAD MODEL PARAMETERS DATABASE CONSTRUCTION UNDER MULTI-SCALE OF TIME

The size and parts of load have been subject to changing formulations. If only one load model revised load data gathered by fault recorder, the revised load model can not reflect time-variant feature of load, which may result in inaccurate or bad presentation for load. This paper classifies power load by time according to load model application scene, time-variant feature of load and statistical synthesis method modelling train of thought, categorizing as follows: short - term load is divided horizontally into morning peak hours, troughs hours, gentle hours, evening peak hours and so on; long-term load can be divided vertically years, seasons, months, workday, weekend and so on. Adopting the above-mentioned categorizing methods, users can choose suited load model as required to do some electric power system calculation for the different profiled scenarios.

3 Gradual Learning Method

All measured load data are needed in parameter estimation using least square, without considering the *chronological* order between measured data. Large quantity measurements request computer has a massive storage capacity. When this process occurs, the measurements are given in chronological sequence. Parameter estimation process can carry on in time order too. The estimates θ can be achieved on the basis of earlier measurements, then estimation results will be revised when new measurements arrive, with it, storage capacity of computer is reduced a lot. This paper identifies model parameters by recursive least square in static load on-line modelling. The identification algorithm comparing with common least square needs little computation, calculates with high speed, doesn't need large amounts of matrix inverse operation, fits in on-line application, and identification results achieved by which are almost the same with calculation achieved by using all measured load data, even better.

The gradual learning method iteration computational equations are given as bellows:

Suppose observation Y is one - dimension, k observations will be achieved by observing k times.

Vectors and matrices expressed by Y_k , H_k , θ and v_k showed as follows:

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, H_k = \begin{bmatrix} h_1(t_1) & h_2(t_1) & \cdots & h_n(t_1) \\ h_1(t_2) & h_2(t_2) & \cdots & h_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(t_k) & h_2(t_k) & \cdots & h_n(t_k) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_k \end{bmatrix}$$

$$v_k = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}, k \geq n.$$

The observing equation can be written in the following matrix form:

$$Y_k = H_k \theta + v_k. \tag{3}$$

The estimate $\hat{\theta}$ can be obtained by the least squares estimation:

$$\hat{\theta}_k = (H_k^T H_k)^{-1} H_k^T Y_k. \tag{4}$$

Suppose

$$P_k = (H_k^T H_k)^{-1}, \tag{5}$$

then

$$\hat{\theta}_k = P_k H_k^T Y_k.$$

If y_{k+1} is the observation for $k+1$ times:

$$y_{k+1} = \theta_1 h_1(t_{k+1}) + \theta_2 h_2(t_{k+1}) + \cdots + \theta_n h_n(t_{k+1}) = h_{k+1} \theta. \tag{7}$$

In the Equation $h_{k+1} = [h_1(t_{k+1}), h_2(t_{k+1}), \dots, h_n(t_{k+1})]$. (7),

Combine Equation (7) and (3):

$$Y_{k+1} = H_{k+1} \theta + v_{k+1}. \tag{8}$$

In (8)

$$Y_{k+1} = \begin{bmatrix} Y_k \\ y_{k+1} \end{bmatrix}, H_{k+1} = \begin{bmatrix} H_k \\ h_{k+1} \end{bmatrix}, v_{k+1} = \begin{bmatrix} v_k \\ v_{k+1} \end{bmatrix}. \tag{9}$$

According to (4),

$$\hat{\theta}_{k+1} = (H_{k+1}^T H_{k+1})^{-1} H_{k+1}^T Y_{k+1}, \tag{10}$$

$$(H_{k+1}^T H_{k+1})^{-1} = \left\{ \begin{bmatrix} H_k^T & \vdots & h_{k+1}^T \end{bmatrix} \begin{bmatrix} H_k \\ \cdots \\ h_{k+1} \end{bmatrix} \right\}^{-1} = \left[H_k^T H_k + h_{k+1}^T h_{k+1} \right]^{-1}. \tag{11}$$

Suppose $P_{k+1} = (H_{k+1}^T H_{k+1})^{-1}$, then

$$P_{k+1} = \left[H_k^T H_k + h_{k+1}^T h_{k+1} \right]^{-1}. \tag{12}$$

Setting by Equation (5),

$$P_k^{-1} = H_k^T H_k. \tag{13}$$

Based on Matrix Inversion Lemma:

$$P_{k+1} = P_k - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k. \tag{14}$$

So Equation (10) can be written as following:

$$\hat{\theta}_{k+1} = \left[P_k - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k \right] \cdot \left[H_k^T Y_k + h_{k+1}^T y_{k+1} \right] = P_k H_k^T Y_k + P_k h_{k+1}^T y_{k+1} - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k \cdot \left[H_k^T Y_k + h_{k+1}^T y_{k+1} \right] = \hat{\theta}_k + P_k h_{k+1}^T y_{k+1} - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} \cdot h_{k+1} P_k H_k^T Y_k - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k h_{k+1}^T y_{k+1} \tag{15}$$

In (15),

$$P_k h_{k+1}^T y_{k+1} = P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} \cdot (h_{k+1} P_k h_{k+1}^T + 1) y_{k+1} = P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k h_{k+1}^T y_{k+1} + P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} y_{k+1} \tag{16}$$

Equation (16) is substituted in Equation (15),

$$\hat{\theta}_{k+1} = \hat{\theta}_k + P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} \cdot \left(y_{k+1} - h_{k+1} \hat{\theta}_k \right). \tag{17}$$

So recursive formula is found based on recursion least square therefrom:

$$\begin{cases} \hat{\theta}_{k+1} = \hat{\theta}_k + P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} \cdot \\ \left(y_{k+1} - h_{k+1} \hat{\theta}_k \right) \\ P_{k+1} = P_k - P_k h_{k+1}^T (h_{k+1} P_k h_{k+1}^T + 1)^{-1} h_{k+1} P_k \end{cases} \quad (18)$$

There are two methods for $\hat{\theta}$ and P to get initialization values. In one way, k initial values can be calculated by Equation (4) and (5) straightly, and set $k=2n$ (n is dimension of model parameters). In another way, $\hat{\theta}(0)$ is arbitrary and $P(0)=\alpha I$ (α is a proper scalar), better and proper initial values can be gotten through recursive steps.

4 Case Simulations

This paper analysis load data of one summer typical workday of output wire from 200KV/100KV transformer simple secondary in Rizhao, data collecting every 5 minutes, typical load data acquired is shown in the Table 1 below. And length of be confined to, the first 10 groups data are displayed only.

TABLE 1 Measured Load Data

Time	U (KV)	P (MW)	Q (MVar)
0:00	116.7	33.64	10.72
0:05	116.9	33.64	10.72
0:10	116.9	32.83	10.59
0:15	116.8	32.97	10.45
0:20	116.6	32.57	10.45
0:25	116.5	32.03	10.45
0:30	116.7	31.90	10.32
0:35	116.6	31.76	10.05
0:40	116.7	31.63	10.19
0:45	116.5	31.90	10.32

Since active and reactive power have almost the same expression throughout, only differ in model parameters, this paper builds static load model based on active power load data, employing polynomial II model structure.

$$P(k) = a_p + b_p U(k) + c_p U^2(k) \quad (19)$$

Error function is shown in (18)

$$J = \sqrt{\sum_{i=1}^n \left(\frac{P(i) - P_m(i)}{P(i)} \right)^2} \quad (20)$$

In the equation, data sampling sites are marked n , P is active power based on load model, P_m is measured active power.

4.1 PARAMETERS IDENTIFIED BASED ON LEAST SQUARES

Static load model is built on one day's load data, identified parameters using all load data based on Least squares are displayed in Table 2. The measured active power is compared with the calculated active power by load model, and illustrated in Figure 1.

TABLE 2 Parameters Identified by The Least Square

a_p	b_p	c_p	Error
0.222755	12.921562	-0.108082	2.645554

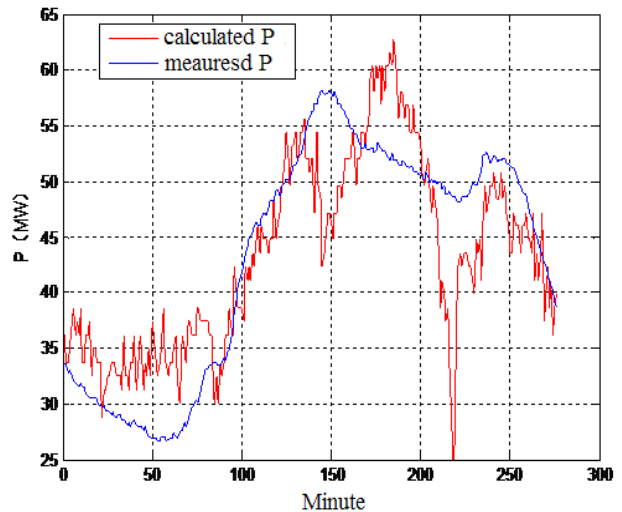


FIGURE 1 Results compared with measured power with calculated power

Table 2 and Figure 1 indicate static load model based on Least squares has weak description, the primary cause is great difference among various times in one day, dispersion and imprecision come to arise when all the load data are used to identify model parameters.

4.2 PARAMETERS IDENTIFIED BASED ON GRADUAL LEARNING METHOD

Load model parameters are identified based on gradual learning method using a whole day's load data, the variation of the parameters identified shown in Figure 2, identified results and error are illustrated in Table 3.

TABLE 3 Identified Results and Error Based on Gradual Learning Method

a_p	b_p	c_p	Error
0.221429	12.921585	-0.108081	2.645554

By comparing Table 2 with Table 3, model parameters identified by gradual learning method is almost the same with that of least squares based on load

data of all time, it is illustrated that static load modelling method based on gradual learning method is feasible. The modelling method just deal with identified results without all storing load data, and is of high efficiency, quick computing speed. Together, the method can be used in on-line static load modelling based on fault recorder conveniently.

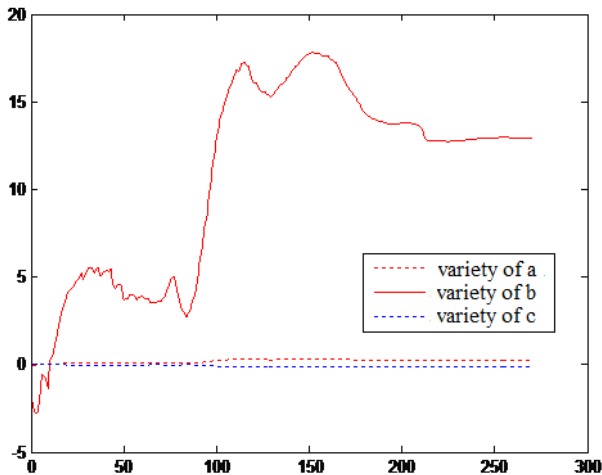


FIGURE 2 The variation of the parameters

As Figure 2 suggests: the trend curve of load model parameters will level off as number of load data increase, as fault recorder operates longer time, model parameters are revised by more measured load data, at the same time, revised parameters are stored only, as time goes by, model parameters become more precisely and static load model is more consistent with practical load.

4.3 STATIC LOAD MODELING BASED ON CLASSIFIED TIME

As is stated above, load model based on all load data in one day is not consistent with practical load very well, because power load has time-varying property. This paper divides load data of one day by the hour, builds static load model and identifies model parameters based on measured load data per hour called hour model, finally, load model for one day is composed by hour models.

The identified parameters results and total error based on the method above are illustrated in Table 4, and comparison of measured active power with calculated active power is shown in Figure 3.

By comparing Table 2 with Table 4, static load model is built on classified time, parameters are identified by recursive least squares, the error between measured active power and calculated active power is smaller than before, precision of load model has been improved a lot. It is understood from Figure 3 the calculated active power is successfully simulated with measured active power, so it seems static load modelling based on classified time is very necessary.

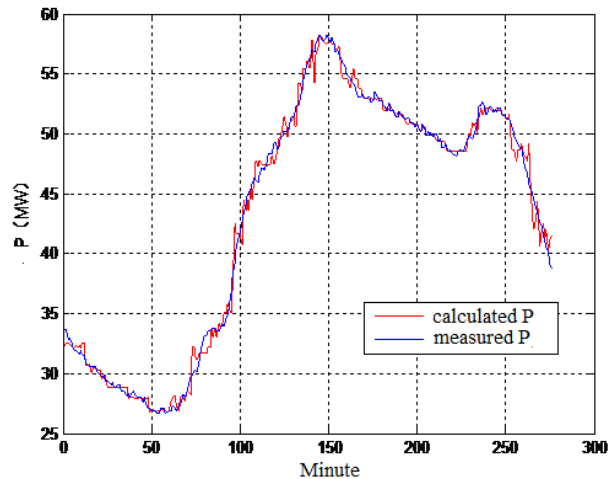


FIGURE 3 Comparison of active power

TABLE 4 Parameters Identified and Total Error Based on Classified Time

hour	a_p	b_p	c_p	Total error
0:00-1:00	-0.01	-0.574	0.007	
1:00-2:00	0.04	2.347	-0.018	
2:00-3:00	0.008	0.496	-0.002	
3:00-4:00	0.012	0.678	-0.004	
4:00-5:00	0.015	0.869	-0.005	
5:00-6:00	0.055	3.206	-0.025	
6:00-7:00	-0.084	-4.874	0.044	
7:00-8:00	0.08	4.676	-0.037	
8:00-9:00	0.178	10.362	-0.086	
9:00-10:00	0.054	3.147	-0.024	
10:00-11:00	0.088	5.072	-0.040	
11:00-12:00	-0.088	-5.093	0.048	0.331306
12:00-13:00	-0.015	-0.870	0.012	
13:00-14:00	-0.055	-3.175	0.032	
14:00-15:00	0.011	0.609	-0.001	
15:00-16:00	0.038	2.157	-0.015	
16:00-17:00	0.040	2.300	-0.016	
17:00-18:00	0.028	1.624	-0.010	
18:00-19:00	0.014	0.800	-0.003	
19:00-20:00	0.069	3.997	-0.031	
20:00-21:00	0.042	2.442	-0.017	
21:00-22:00	0.107	6.208	-0.050	
22:00-23:00	0.093	5.404	-0.043	

5 Conclusions

This paper proposes a static load modelling method based on real-time data collected by fault recorder. Modern power fault recorder can gather and store plenty of real-time load data, valid load data files is set for on-line static load modelling. All data will be classified by the time, and model parameters database will be built correspond to each classification. Model parameters are identified by recursive least squares, at the same time database is renewed. Simulation results proved the static load modelling method is validity and feasibility.

References

- [1] He Renmu, Wei Xiaoming, Han Minxiao 1996 Power System Dynamic Load Modeling Based on the Measurements in the Field *Proceedings of the CSEE* **16**(3) 151-154
- [2] Asif Islam, Samiur Rahman Hasib, Md. Shariful Islam 2013 Short Term Electricity Demand Forecasting of an Isolated Area using Two Different Approach *Journal of Power Technologies* **93**(4) 185-193
- [3] Zhang Linli, Zhou Wen 1999 The synthesis of dynamic load characteristic *Proceedings of the CSEE* **19**(9) 36-45
- [4] Shi Jinghai, He Renmu 2004 Load time-variantion study in dynamic load modelling *Proceedings of the CSEE* **24**(4) 85-90
- [5] IEEE Task Force on Load Representation for Dynamic Performance 1995 Standard load models for power flow and dynamic performance simulation *IEEE Trans on Power Systems* **10**(3) 1302-13
- [6] Li Xinran, Chen Yuanxin, Jiang Tiezheng 2000 Power load model and its modeling method for voltage stability analysis *Proceedings of the Chinese Society of Universities for Electric Power System and Automation* **12**(6) 9-13 (In Chinese)
- [7] Li Peiqiang, Li Xinran, Chen Feng 2003 Application of fuzzy clustering in component-based modelling approach *Power Automation Equipment* **23**(5) 42-45
- [8] M K Pal 1995 Assessment of Corrective Measures for Voltage Stability Considering Loads Dynamics *Electrical Power & Energy Systems* **17**(5) 325-34
- [9] Ju Ping, Ma Daqiang 2008 *Power system load modelling* China Electric Power Press: Beijing (In Chinese)
- [10] Ju Ping, Jin Yan, Wu Feng 2004 Studies on classification and synthesis of composite dynamic loads *Automation of Electric Power Systems* **28**(1) 64-68
- [11] Li Xinran, Lin Shunjiang, Liu Yanghual 2006 A new classification method for aggregate load dynamic characteristics based on field measured response *Proceedings of the CSEE* **26**(8) 39-44
- [12] Lin Shunjiang, Li Xinran, LI Peiqiang 2006 A novel direct synthesis method for aggregate load dynamic characteristics based on field measured response *Proceedings of the CSEE* **26**(21) 36-42
- [13] Byoung-Kon Choi, Hsiao-Dong Chiang, Yin hong Li 2006 Measurement-Based Dynamic Load: Derivation, Comparison and Validation *IEEE Transaction on Power Systems* **21**(3) 1276-83

Authors



Guoping Shi, born in November, 1977, Lixia District, Jinan City, P.R. China

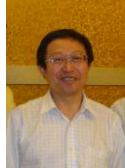
Current position, grades: lecturer of school of Information and Electrical Engineering, Shandong Jianzhu University, China.

University studies: B.Sc. from Shandong University in China. M.Sc. from Shandong University in China.

Scientific interest: Power system control, Power load modelling.

Publications: more than 7 papers published in various journals.

Experience: teaching experience of 10 years, 6 scientific research projects.



Jun Liang, born in August, 1956, Lixia District, Jinan City, P.R. China

Current position, grades: Professor of School of Electrical Engineering, Shandong University, China.

University studies: B.Sc. from Shandong University in China. M.Sc. from Shandong University in China.

Scientific interest: Power system control, Power load modelling.

Publications: more than 100 papers published in various journals.

Experience: teaching experience of 30 years, 10 scientific research projects.

The scroll flow and torque prediction with the wavelet neural network optimized by PSOA and BP

Ying Kong^{1*}, Xiao guang Chu²

¹*School of Medical Information and Technology, JiNing Medical University, Rizhao 276826, China*

²*School of Electrical information automation, QuFu Normal University, Rizhao 276826, China*

Received 1 May 2014, www.tsi.lv

Abstract

A new Compressed Air Energy Storage(CAES) with scroll was proposed to promote the storage efficiency, which can be acquired by the scroll efficiency tracking control with the timely evaluation, but the flow and torque is not easy acquired because the sensors possessed the merits of high price, lower life-span and subjection to the disturbance, so a torque and flow prediction algorithm based on Wavelet Neural Network (WNN) is proposed adopting a hybrid learning algorithm combining Particle Swarm Optimization (PSO) with BP. Through the comparison between predictive and the experimental data and the scroll efficiency experiment, the proposed prediction method is validated and can be successfully used to improve Pneumatic conversion efficiency.

Keywords: Compressed Air Energy Storage, Scroll, Particle Swarm Optimization (PSO), Wavelet Neural Network

1 Introduction

With the increasing depletion of conventional energy and the worsening environmental pollution, wind energy and solar energy have become the new driving force for the global economic growth, but the intermittent and the fluctuation seriously affect the power quality. Many researches show that the storage is the effective way to solve the above the problem.

Due to long span life and non-pollution, CAES is a promising storage way to stabilize the power fluctuation [1], however the lower storage efficiency seriously limits the real application of the CAES. Efficiency tracking is an effective method to enhance storage efficiency, but must real-timely evaluate the scroll performance by the flow and torque, but the flow sensor is vulnerable to environmental disturbance making the measured data deviate from the actual value. Furthermore, adding flow sensor and corresponding filter increases pressure loss and reduces storage efficiency. In addition, Surge is harmful and easily to occur at smaller flow, so it is important to real-timely monitor the flow and torque. In addition, torque sensor is all expensive and easy to be destroyed. Therefore, it is not advisable to measure the flow and torque by sensors, which make it necessary to find a new effective approach to acquire the flow and torque.

As the literatures [2, 3] described, scroll models are so complicated that could not be used to predict the flow and torque, which all modelled for performance analysis. In view of effective nonlinear modelling and prediction ability of WNN [4, 5], WNN can be applied to predict the scroll flow and torque. WNN originated from the

decomposition in signal processing, which has become more popular lately [5, 6]. Zhang [6] proposed a new notation of wavelet network as an alternative to feed-forward neural networks for approximating arbitrary nonlinear functions based on wavelet transform theory. However, the architecture and the learning algorithm of WNN severely affect the ability of predicting and modelling, so many researchers have been studying on the parameters optimization of WNN, but all were focused on adjusting the initial parameters. Kuok proposed the PSOA for optimizing feed-forward neural networks, which acquired excellent performance because of rapid global search ability of PSOA [7]. Meanwhile Back Propagation Network (BP) owns the faster local optimization ability, which can compensate little local optimization ability of the PSOA. For above reason, the WNN optimized by the PSOA and BP is proposed for predicting the scroll flow and torque [6-8].

2 Scroll Performance Analysis

As shown in Figure 1, the new compress air storage system with scroll is proposed to solve wind power fluctuations at front of the generator. Scroll can be operating in the compressor mode and expander mode. The scroll can be running in the compressor mode when wind capture power is more than the load power, and the excess power will be stored in the storage vessel.

While the wind power could not supply for the load, the scroll will be working on the expander mode to supply the deficient power. High efficiency of the scroll is important for the real application of the compressed air storage system, but the scroll efficiency is affected by the

* *Corresponding author's* e-mail: kongyuchu@163.com

speed and the discharge pressure, so the maximum efficiency tracking is important to make the scroll run the highest efficiency under any conditions.

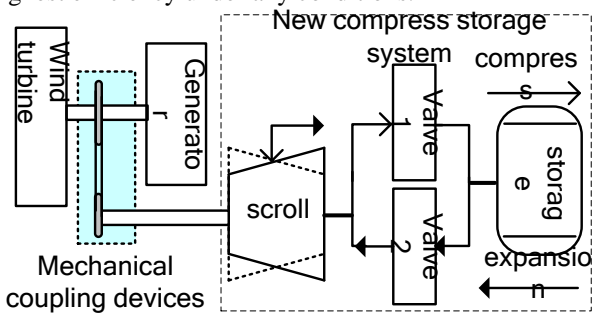


FIGURE 1 New Compressed Air Energy Storage system with scroll

But the scroll maximum efficiency tracking must be finished by the SC efficiency evaluation timely, through which to optimize control scroll. The scroll efficiency can be evaluated by Equation (1).

$$\eta_{scroll} = \frac{p_d Q_d \ln(\frac{p_d}{p_{atm}})}{T_{in} \omega}, \tag{1}$$

where p_d is the discharge pressure; Q_d is the discharge flow; T_{in} is driving torque, and ω is angular speed.

As the Figure 2 described, scroll machine consists of two intermeshed identical scrolls, which form many closed chambers, and chamber air will be pressuring successively with the Chambers volumes change gradually. When orbit angle revolves to the discharge angle, the compressed air can be discharged to the storage. Driving Torque has strong relation with the discharge pressure p_d . In view of the fixed volumetric machine of the scroll, scroll flow depends on the speed and the inlet volume. Hence the speed and storage pressure are the key influence factor on flow and torque, which will be taken as the principal input for the prediction.

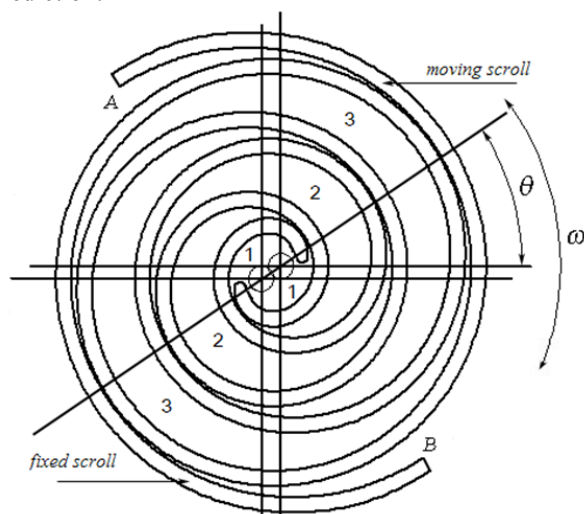


FIGURE 2 Scroll machine structure

3 Scroll flow and torque prediction algorithm

The WNN optimized by the hybrid learning algorithm combining the PSO with the BP algorithm is applied to predict the flow and torque of the scroll. For sake of quickly acquiring optimal result, this algorithm uses the PSO to do global search in the beginning, and then uses the BP to do local search around the global optimum.

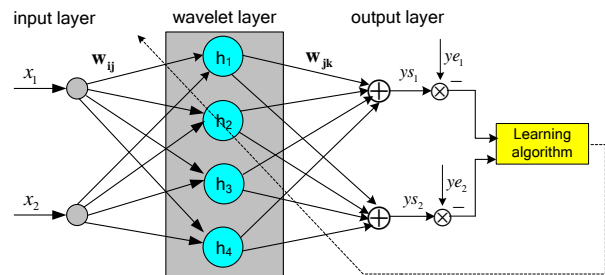


FIGURE 3 Three-layer WNN structure

3.1 THE STRUCTURE OF THE WNN

As the key factor influencing the flow and torque of the scroll, the speed and discharge pressure was taken as the input variable (x_1 and x_2), while the flow and torque acted as the output variable (ys_1 and ys_2). The structure of the WNN model is shown in Figure 3. It is designed as a three-layer structure of 2-4-2, which comprised of an input layer, a wavelet layer, and an output layer. For the discrete wavelet transform, the mother wavelet $\varphi(x)$ describes the dilation a and the translation b as follow:

$$\varphi_{a,b}(x) = \varphi\left(\frac{x-b}{a}\right). \tag{2}$$

In this paper, we use the Morlet wavelet as the wavelet function which can be controlled and expressed as follow:

$$\varphi(t) = \cos\left(\frac{7}{4}t\right) e^{-\frac{t^2}{2}} \tag{3}$$

Therefore, the activation function of the j^{th} wavelet node connected with the two input data is represented as:

$$h(j) = \cos\left(\frac{7}{4} \sum_{i=1}^2 \frac{w_{ij}x_i - b_j}{a_j}\right) e^{-\frac{1}{2} \left(\sum_{i=1}^2 \frac{w_{ij}x_i - b_j}{a_j}\right)^2} \quad j = 1 \sim 4 \tag{4}$$

According to the theory of multi-resolution analysis, output can be regarded as a linear combination of wavelets, so it can be expressed as:

$$ys = \sum_{j=1}^4 w_{jk} h(j), \tag{5}$$

where w_{ij} and w_{jk} are the input-wavelet layer and wavelet-output layer adjustable weighting parameters, a_j and b_j are the dilation and translation factors, which all are optimized to improve the performance of the WNN.

The fitness function of the k^{th} training sample can be defined as follows:

$$fitness(k) = \frac{1}{2} \sum_{i=1}^2 (ys_i(k) - ye_i(k))^2, \tag{6}$$

here i is the number of the output node; k is number of the sampling data; $ys_i(k)$ is the model prediction output, while $ye_i(k)$ is the actual output, $e(k)$ is the learning error.

3.2 THE PARAMETER OPTIMIZATION ALGORITHM FOR THE WNN

Based on these considerations of the relative merits analysis of PSO and BP, the project adopts a hybrid algorithm combing with PSO and BP to optimize parameters of WNN for flow and torque prediction. As shown in Figure 4, the WNN key parameters is optimized including three steps, I) PSO algorithm is used to find the optimum region for the initialization of the BP algorithm; II) BP algorithm is applied to find turn optimum solution; III) optimal result is determined by the comparison of the PSO and BP result.

3.2.1 PSO optimization process

This project, PSO algorithm is initialized with a group of random particles including weights, translation factor and dilation factor. Each particle can be represented as:

Particle(i)=[w_{11} w_{12} w_{13} w_{14} w_{21} w_{22} w_{23} w_{24} w_{31} w_{32} w_{33}
 w_{34} w_{41} w_{42} w_{43} w_{44} a b],
 Particles matrix = [Particle(1); Particle(2); ... Particle(n)] ,

here m is the total particles, and this project set the particles as 200.

Every particle holds its velocity $v_{id}(t)$ and position $x_{id}(t)$, which can be dynamically adjusted by

$$v_{id}(t+1) = wv_{id}(t) + c_1 \cdot r_1 \cdot [pb_{id}(t) - x_{id}(t)] + c_2 \cdot r_2 \cdot [pg_d(t) - x_{id}(t)] \tag{7}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{8}$$

$$w(t) = w_0 - (w_0 - w_{end})(t / maxgen1)^2 \tag{9}$$

where, $pb_{id}(t)$ is the best-found location of i^{th} particle at the time; $pg_d(t)$ is the global best position among all particles up to time; c_1 and c_2 are two positive constants, while r_1 and r_2 are two random parameters; w is the inertia weight to balance between global and local explorations, which can be adjusted according to the Eq. (9), and w_0 and w_{end} are the initial and the end inertia weights; $maxgen1$ is the maximal iterative times of the PSO, and this project set it as 500.

As shown in Fig.4, PSO optimization is repeated until the convergence criterion or stop conditions are satisfied, which can be set as $fitness(k) \leq 0.001$ and $maxgen1=500$. Or else, the velocity, position and w are renewed repeatedly, and then performance criterion is evaluated according to Equation (6). Then, best position and group best position can be updated repeatedly until the predefined convergence criterion is satisfied, which can be set as the initialisation for the BP algorithm.

3.2.2 BP optimization process

BP will start to search the optimum around the pg that the PSO acquired until the stop conditions are met. The optimal stop condition sets as $maxgen2=1500$, and the best optimal solution can be adjusted by the Equation (10) [9] repeatedly.

$$\begin{cases} W_{ij}(t+1) = W_{ij}(t) + \eta \Delta W_{ij}(t) = W_{ij}(t) - \frac{\eta}{2} \sum_{h=1}^2 \frac{\partial ys_h}{\partial W_{ij}} e \\ W_{jk}(t+1) = W_{jk}(t) + \eta \Delta W_{jk}(t) = W_{jk}(t) - \frac{\eta}{2} \sum_{h=1}^2 \frac{\partial ys_h}{\partial W_{jk}} e \\ a(t+1) = a(t) + \eta \Delta a(t) = a(t) - \frac{\eta}{2} \sum_{h=1}^2 \frac{\partial ys_h}{\partial a} e \\ b(t+1) = b(t) + \eta \Delta b(t) = b(t) - \frac{\eta}{2} \sum_{h=1}^2 \frac{\partial ys_h}{\partial b} e \end{cases} \tag{10}$$

where η is the learning rate, and the $e = ys(t) - ye(t)$.

3.2.3 Optimal result decision process

BP algorithm searches the optimum around Pg for some generations (1500). If the stop condition is met, the optimization must be decided by the comparison with PSO and BP result, and the smallest can be transmitted to WNN.

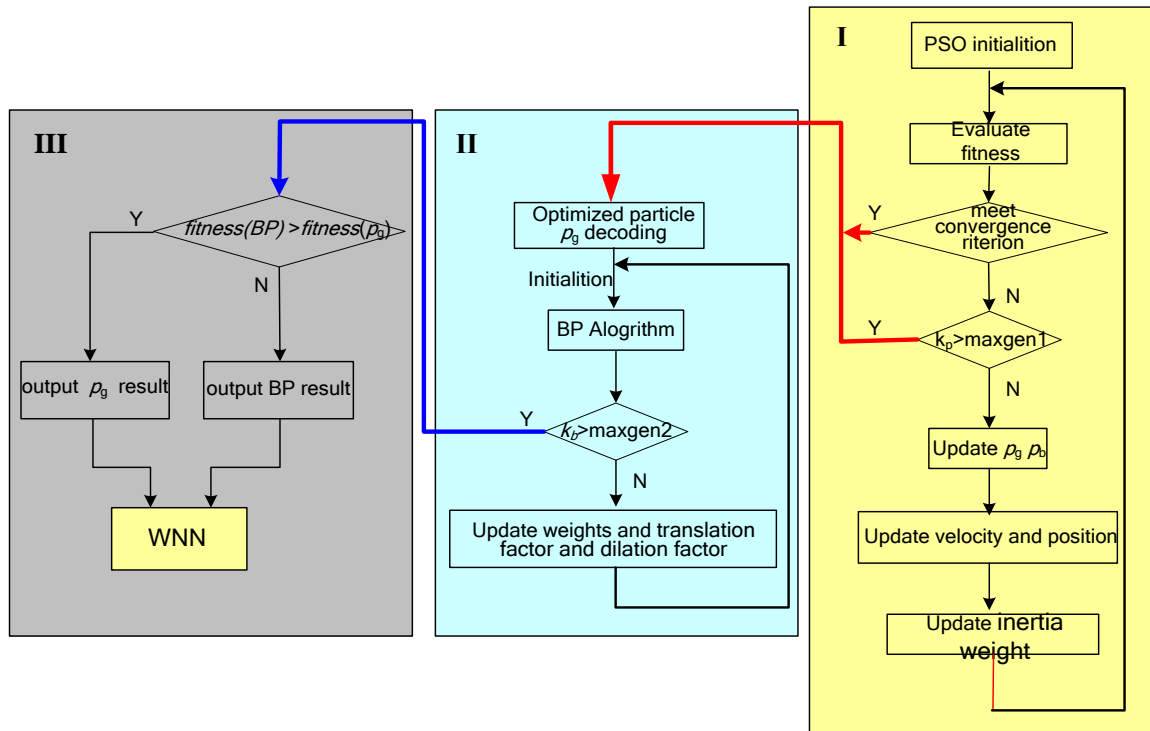


FIGURE 4 WNN key parameters optimization principle

Flow sensor Pressure sensor scroll Torque and speed sensor motor

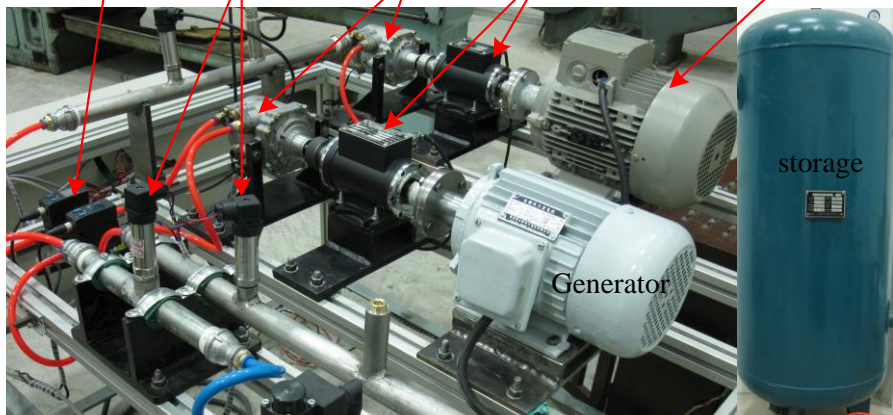


FIGURE 5 New CAES with scroll experiment test rig

4 Experiment Result Analysis

As shown in Figure 5, the new CAES experiment test rig is set up including 4kW Siemens motor, scroll machine(WHX-86, discharge pressure: 0-600kpa), storage vessel(1.5m³), scroll flow and torque. Measuring and acquiring system. Multi-speed scroll experiment are performed by motor, and then 3580 groups data including speed, discharge pressure, flow and torque were collected, and 3000 groups were applied to train and optimize the WNN key parameters, while the rest were used to validate the model identified.

As shown in Figure 6, average fitness of the flow and torque are respectively 0.01 and 0.013. The model prediction output has been successful in approaching the

experiment torque and flow, which validated the WNN optimized by hybrid training algorithm efficient.

As shown in Figure 7, scroll efficiency tracking is realized from the timely efficiency evaluation by the predicted flow and torque. With the increase of discharge pressure, the scroll is running with the highest efficiency all the time. When discharge pressure less than the designed pressure, the efficiency tracking is acquired by the elevate speed to accelerate the efficiency enhancement corresponding to the time from 0s to 1500s; while the time from 1500s to 3500s, the discharge pressure is larger than the designed pressure, the efficiency tracking is realized through the raising speed to prevent the efficiency reduction. The efficiency tracking experiment further validated the prediction algorithm, which can be directly used to the CAES system.

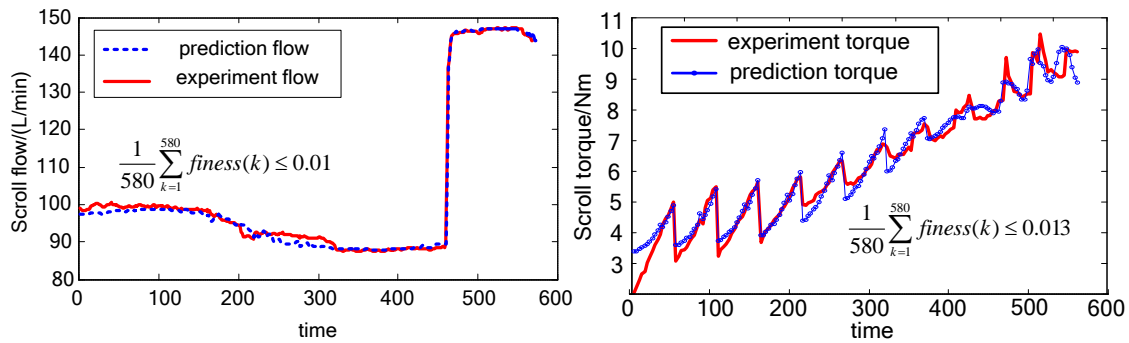


FIGURE 6 The comparison with the prediction and experiment result

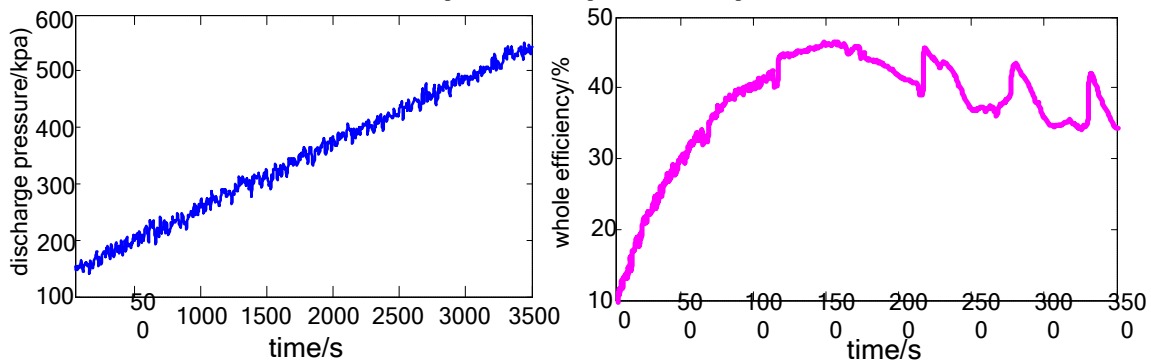


FIGURE 7 Scroll maximum efficiency tracking with the flow and torque prediction

5 Conclusion

This paper proposed a flow and torque prediction algorithm adopting WNN optimized by the PSO and BP. Key parameters of WNN is trained by the multi-speed experiment data. With comparison of the prediction and experiment data and the scroll efficiency tracking experiment, the prediction algorithm of the flow and

torque is validated, which can be successfully to improve the Pneumatic conversion efficiency.

Acknowledgements

The authors would like to thank QF Normal University Foundation (2013kj0009) and JiNing Medical University Foundation (JY2013KJ031).

References

- [1] Arsie I, Marano, Nappi V G, Rizzo G 2005 A model of a hybrid power plant with wind turbines and compressed air energy storage *Proc of PWR 2005 ASME Power Chicago Illinois* 2005-50187
- [2] Wang Baolong, Li Xianting, Shi Wenxing 2005 A general geometrical model of scroll compressors based on discretional initial angles of involute *International Journal of Refrigeration* 28(6) 958-66
- [3] Chen Yu, Halm N P, Groll E A, Braun J E 2002 Mathematical modeling of scroll compressors part I: compression process modeling *International Journal of Refrigeration* 25(6) 731-50
- [4] Lin Cheng-Jian, Chen Cheng-Hung, Lee Chi-Yung 2004 A self-adaptive quantum radial basis function network for classification *Applications Proc. of 2004 IEEE International Joint Conference on Neural Networks* 4 3263-8
- [5] Huang D S 1996 *Systematic Theory of Neural Networks for Pattern Recognition* Publishing House of Electronic Industry of China, Beijing
- [6] Zhang Jing-Ru 2007 A hybrid particle swarm optimization bp algorithm for feed forward neural network Training *Applied Mathematics and Computation* 185(2) 1026-37
- [7] Kuok K K 2010 Particle swarm optimization feed forward neural network for modeling runoff *Proc. of IEEE Int. Conf. on System Int Environ Sci Tech* 7(1) 67-78
- [8] Zhang Chunkai, Shao Huihe, Li Yu 2000 Particle swarm optimization for evolving artificial neural network[C] *Proc. of IEEE Int. Conf. on System, Man, and Cybernetics* 4 2487-90
- [9] van Ooyen A, Nienhuis B 1992 Improving the convergence of the back-propagation algorithm *Neural Networks* 5(3) 465-71

Authors	
	<p>Ying Kong, born in 1975, Shandong, China</p> <p>Current position, grades: lecturer of computer science at School of Medical Information and Technology, JiNing Medical University University studies: master degree of engineering in control theory and control engineering from Shandong University of Science and Technology, China, in 2006. Scientific interest: complex control algorithm, new energy power generation and compressed gas energy storage.</p>
	<p>Xiao guang Chu, born in 1975, Shandong, China</p> <p>Current position, grades: lecturer of Technology School of Technology University studies: M.S in 2007, the doctor degree from Shandong university in 2012 Experience: maglev technique, new energy power generation and compressed gas energy storage.</p>

A study on the acoustic emission characteristics of the coal rock on different bedding direction

Xiaohui Liu^{1, 2*}, Xiaoping Zhao³, Jianfeng Liu²

¹College of Energy and Environment, Xihua University, Chengdu, Sichuan, China, 610039

²College of Water Resources and Hydropower, Sichuan University, Chengdu, Sichuan, China, 610065

³Chengdu Engineering Corporation, Power China, Chengdu, Sichuan, China, 610072

Received 6 June 2014, www.tsi.lv

Abstract

The rock mechanics servo system (MTS815) and the acoustic emission testing system (PCI-2) were selected in this paper to conduct the uniaxial compression test and the acoustic emission test from two different direction of the coal and to clarify the acoustic emission feature during the process compression deformation of the coal. The results show that there is an obvious difference on the characteristic parameter of acoustic emission and the spatial distribution during the process of uniaxial compression, which is from different directions. On the initial compression phase, the stress growth rate of the coal from the vertical direction is greater than that from parallel bedding coal and the acoustic parameters. Acoustic emission ringing, energy and event count rate are less than the coal rocks from parallel bedding. The acoustic emission count will increase with arise stress, the coal from the parallel direction is more stable than the coal from the vertical direction; the coal acoustic emission has significantly reduced which from the vertical direction when after the peak stress and there is a little change with the parallel coal rock; The acoustic emission event of the coal which is from the vertical direction and the parallel coal usually concentrates in the lower part of the coal.

Keywords: Coal Rock, Acoustic Emission, Spatial Distribution of Events, Bedding Direction, AE Characteristic Parameters

1 Introduction

In the process of the mining, the stress of the coal may has changed due to the original stress balanced state was destroyed. The phenomenon which can rapidly release the strain energy is called acoustic emission, AE. The acoustic emission technology is an acoustic method to analyse the mechanic of the rock.

The acoustic technology can reflect the internal defects by means of the acoustic emission signal with the different conditions and infer the behaviour variation within the coal rock. These technologies can also back analyse the failure mechanism of the coal rock. For recent years, with increasingly testified the exploitation intensity and the depth of the coal in China, the acoustic technology will be more and more widely applied. The research on acoustic emission characteristics has made a lot of achievements in recent years. Yang Yong-jie discovered that the maximum Lyapunov index for the energy counting rate which can be used for predicting the coal fracture [1]. Zuo Jian-ping has combined the single rock, the single coal and the coal rock together, the single axis acoustic emission testing was conducted on this compound in order to analyse the temporal and spatial evolution mechanism during the compound crack process [2]. Ning Chao found that there is a variation curve relationship between the stress of the coal and the acoustic emission parameters [3]. Qin Hu found that the

moisture content in the coal rock will have an obvious influence on the intensity and the acoustic emission of the coal rock [4]. Voznesenskii found that the relative coefficient of the acoustic emission is a constant [5]. Majewska considered that there are some differences with the emission and the strain characteristics during the process of carbon dioxide adsorption [6]. Cao Shu-gang found that the acoustic emission characteristic may be different under the impact of the uniaxial compression and confining pressure [7]. Ai Ting has located the acoustic emission under the different confining pressures and considered that the acoustic phenomenon mainly comes from mid-prophase [8]. Gao Bao-bin has researched the acoustic emission and found out the comparative rules of the fractal characteristics between the acoustic emission count and the cumulative count [9]. Su Cheng-dong found that the deformation and destruction process of the coal might be different under the different stress path [10].

There is a certain relationship between the heterogeneity and the mechanical property of the coal rock [11]. As mentioned before, the researches on the acoustic emission characteristic were mainly concentrated on the uniaxial compression and the confining pressure of the coal. However, few researches focused on the acoustic emission characteristic concerning the bedding direction. The rock mechanics servo system (MTS815) and the acoustic emission system (PCI-2) were selected to test the uniaxial compression. Furthermore, we will

*Corresponding author e-mail: lxh_1001@tom.com

conduct the synchronization sound emission research and analyze the acoustic emission feature in the process of the deformation and fracture of the coal.

2 The basic theory of the coal rock acoustic emission

In continuous damage mechanics, the damage can be defined as variable D and calculated as follow [12]:

$$D = \frac{S_n}{S}, \tag{1}$$

where S is the material section surface without damage; S_n is the damage area; D is the damage variables, where $0 \leq D \leq 1$.

It is considered that the material is entirely damaged when the variable D equals to 1, the material is in completely non-invasive state when D equals to 0. The coal material is the infinitesimal which has a defective, the destroy essence of the coal is the inherent defect which suffers a stretch and extend, like the microcosmic breakage. So the flaw may be activated when the stress is greater than the limit fracture strength and releases the strain elastic energy to motivate the acoustic emission, two-parameter Weibull distribution is often used in the coal defective distribution:

$$n(\varepsilon) = k\varepsilon^m; \tag{2}$$

$$n'(\varepsilon) = km\varepsilon^{m-1}, \tag{3}$$

where strain $n(\varepsilon)$ is the number of activated flaw; The constants k and m representative the material property of the breakage; $n'(\varepsilon)$ is the gradient rate between the flaw and the strain.

When the strain adds an increment ($d\varepsilon$), the number of defects, which join the breakage activity is:

$$dn = n'(\varepsilon)d\varepsilon. \tag{4}$$

Due to the damage, the stress in the materials was released, it is assumed that there is a corresponding relationship between the active damage destroys and the acoustic emission count, where the acoustic emission counts can be expressed as follow:

$$dN = (1-D)n'(\varepsilon)d\varepsilon. \tag{5}$$

We add the formula (3) into (5) and get:

$$dN = km(1-D)\varepsilon^{m-1}d\varepsilon. \tag{6}$$

The random factor $r(\varepsilon)$ was considered the flaw distribution of the coal rock, the total acoustic emission count (N) can be elicited as formula (7):

$$N = r(\varepsilon) \int_{\varepsilon_0}^{\varepsilon} km(1-D)\varepsilon^{m-1}d\varepsilon, \tag{7}$$

where the strain of the materials initial damage is ε_0 , $r(\varepsilon)$ is the random factor and the value can range from 0 to 1.

The sound emission gradient is:

$$N'(t) = \frac{dN}{dt} = r(\varepsilon)km(1-D)\varepsilon^{m-1} \frac{d\varepsilon}{dt}. \tag{8}$$

The acoustic emission of the uniaxial compression and damaged process can be described as formula (7) and formula (8) [13].

We conclude from the formula that the value of the acoustic emission and the acoustic gradient depends on the damage factor, the transient strain and the strain rate. The value also has close relations to the defectiveness amount, dimension, bedding distribution and the homogeneity.

3 The preparation of the coal rock

The coal sample is from Furongbaijiao colliery, the fault age in the mine has fully development and has a clearly stratum distribution. Accordance with the rule “the physical and mechanical property assay method of the coal and rock (GB/T 23561.7-2009)”, the coal rock can be made up a standard sample ($\phi 50 \times L100mm$) according to the vertical and the parallel direction, shown in Figure 1.

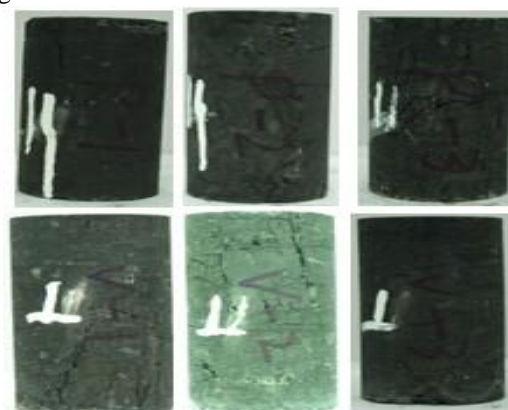


FIGURE 1 Pictures of coal rock

During the process of sample preparation, we must guarantee the two end face of the sample is less than the 0.05mm and the axial direction is less than 0.25. The basic parameter of the coal rock is described in Table 1.

TABLE 1 The basis parameters of the coal rock

Bedding direction	Coal Number	Diameter/mm	Height /mm	Volume /cm ³	Weight/g	Density kg/m ³	Crush intensity/MPa	Elasticity modulus/MPa
Parallel	P-1	50.33	81.86	162.78	233.50	1434.47	6.571	1914
	P-2	50.07	100.18	197.15	304.00	1541.94	8.986	1201
	P-3	50.43	90.19	180.06	258.54	1435.89	12.596	1710
	mean	50.28	90.74	180.00	265.35	1470.77	9.384	1608
Vertical	V-1	50.06	101.93	200.52	293.11	1461.76	14.618	1761
	V-2	49.86	100.32	195.78	291.46	1488.73	9.582	1413
	V-3	49.72	100.09	194.23	293.77	1512.46	15.624	1921
	mean	49.88	100.78	196.84	292.78	1487.65	13.275	1698

4 Experimental facilities

The experimental facilities are provided by the geotechnical engineering key laboratory, Si Chuan University. The mechanics-practical system of the rock (MTS815 Flex Text GT) is used in this research, the maximum axial direction is 4600kN and the crosswise range is from -4 to +4mm, the axial direction is from 0 to 100mm.

In order to guarantee the synchronism of the system, the acoustic emission and the loading systems are selected to monitor the experiment. The PCI-2 acoustic emission system is provided by PAC Company and this system has several advantages which include the ultrafast processing speed, the low threshold value and the reliable stability. The main parameter of the sound emission will be recorded automatically in the process of the experiment. The experiment equipment is shown in Figure 2.

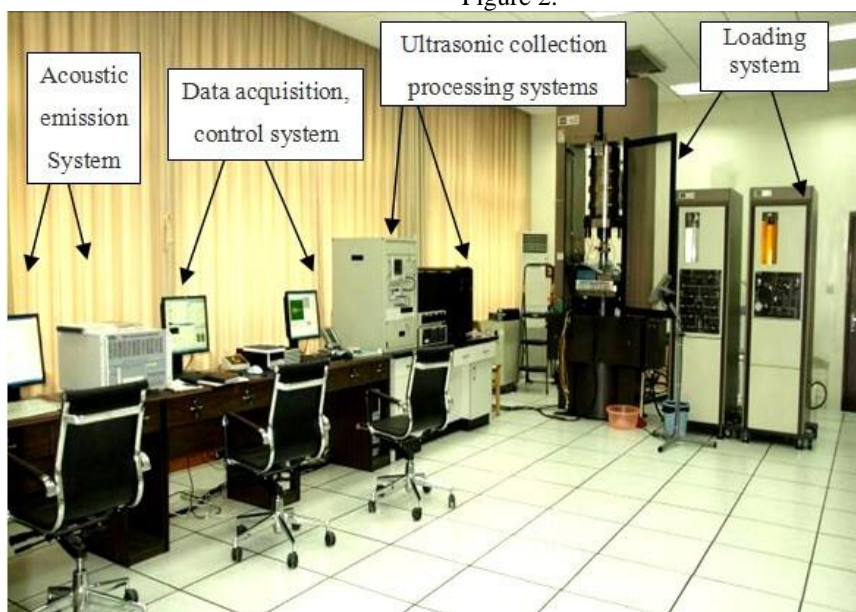


FIGURE 2 MTS hydraulic servo-control testing machine

5 Results and analysis

5.1 THE ACOUSTIC EMISSION ANALYSIS

After the uniaxial compression test, we find the acoustic emission counting rate should correspond to each time quantum, the homotaxial press process is considered in two directions (the vertical and the parallel direction). The relationship between the acoustic emission parameters and the time are shown in Figure 3 and Figure 4.

1) When comparing the two figures, we find that the growth rate of the bedding coal from the vertical direction is larger than that from the parallel direction at the initial

loading stage; The stress value of the coal from the vertical direction will appear minimum value at the 1500s before arise. The stress value of the coal bedding in the parallel direction will decline around 800s and before decline.

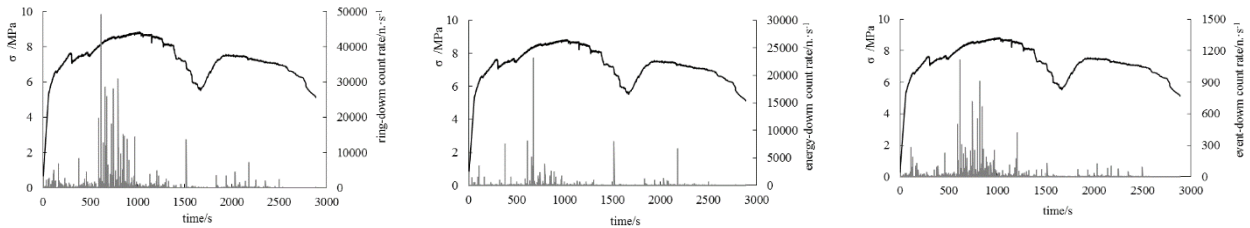
2) Figures 3a and 4a represents the variable curve, the curve can describe the ringing counting rate of the coal which can be changed over time. From the figure we can see that the ringing counting rate from the parallel direction is greater than that from the vertical direction and this may explain that the acoustic emission feature is determined by the layer property; with the increasing the stress value, the ringing counting rate can reach maximum value when in vicinity of the peak value. The ringing counting rate of the vertical direction ranges from

600s to 1000s and belongs to fracture [14]. But the ringing count rate is uniform when at the process of the rock pressure and belongs to stable type.

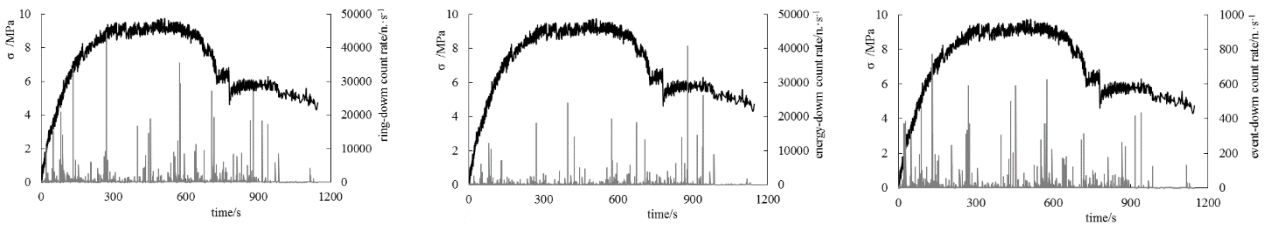
3) Figures 3b and 4b represent the energy counting rate curve over time. As can be seen from the figure, the maximum value of the energy counting rate from the vertical direction appears nearby peak value, but the maximum from the parallel present to late peak and has a more uniformity distribution than the vertical's. This

phenomenon clarifies that the acoustic emission may be different at the different direction of the coal.

4) Figures 3c and 4c represent that the incident counting rate has changed with the time. Figures 3c and 4c have similarities with Figures 3a and 4a correspondingly, the maximum counting rate of the coal from the vertical direction equals to the raining counting rate, but the incident counting rate is different.



(a) Ring-down count rate (b) Energy count rate (c) Event count rate
 FIGURE 3 AE characteristic parameters and stress-time curve for the vertical bedding coal rock (P-2)



(a) Ring-down count rate (b) Energy count rate (c) Event count rate
 FIGURE 4 AE characteristic parameters and stress-time curve for the parallel bedding coal rock (V-2)

5.2 THE DISTRIBUTION OF THE ACOUSTIC EMISSION

Figures 5 and 6 represent the acoustic emission from the vertical and the parallel direction respectively. Figures 5a and 6a are spatial incident distribution map in the increment stage of the AE and Figures 5b and 6b are the stress accumulative incident of the AE.

1) From Figures 5a and 6a we can see that there is a larger difference between the spatial distribution of the acoustic emission from the vertical direction and the parallel direction. In the initial stage of the press, the micro-crack and the micro-structure may cause the acoustic emission incident. The acoustic emission incident appears on the coal sample uniformly, but the firing point from the parallel direction usually focuses on the middle section of the coal; the initial micro-crack will close along with the stress increasingly grown. The acoustic emission incident has reduced to the range from 40% to 50% and formed a failure mode. The acoustic emission incident will drop off with the increasing stress.

2) From Figures 5b and 6b we find that along with the increasing stress, the coal rock will experience from initial crack closure to new micro-crack extend, cut-through and then destroy. The acoustic emission of the coal from the vertical direction spreads all over the coal rock, but the acoustic emission from the parallel direction

mainly concentrates upon the bottom section of the coal and destruction of the coal is mainly the separation.

6 Conclusions

1) This paper applies the coal rock damage theory and the inherent defect distribution of the coal rock to clarify that there is a relationship between the acoustic emission characteristics and the bedding of the coal.

2) There is a big difference between the acoustic emission in the uniaxial compression test over different direction and the incident spatial distribution. The results prove that it is feasible using the acoustic emission technology to analyse the mechanical properties with different directions.

3) In the initial stress stage, the stress of coal from the vertical direction has increased rapidly and the coal has a lower acoustic emission; Along with the increasing stress, the acoustic emission level has increased.

4) In the initial stress stage of the parallel direction, the stress growth rate from the vertical direction changes slowly, but the sound emission, energy and the incident counting rate are high. The acoustic emission events usually distribute in the lower part of the coal rock and have an obvious acoustic emission concentration area, which corresponds to split.

Acknowledgements

The study was supported by the national key basic research development plan (937) funded project (2011CB201201, 2010CB226802, 2010CB732005).

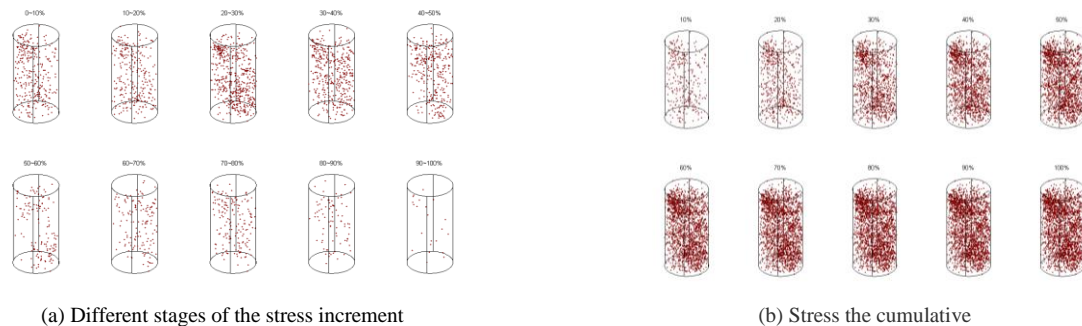


FIGURE 5 AE event space distribution for the vertical bedding coal rock (P-2)

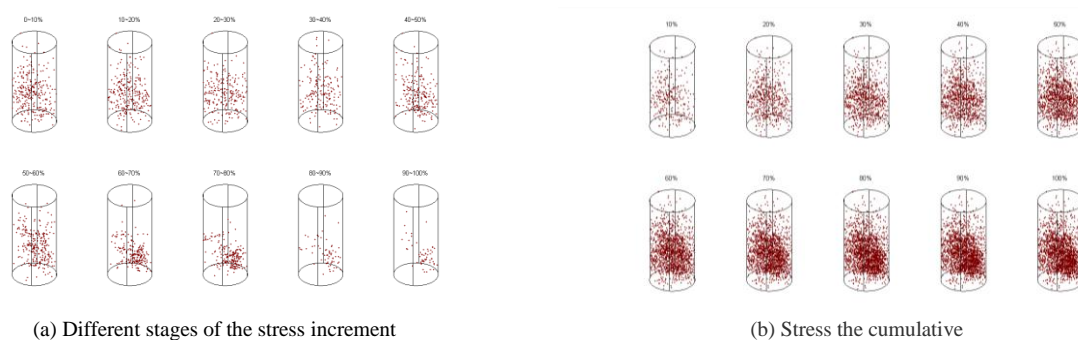





FIGURE 6 AE event space distribution for the parallel bedding coal rock (V-2)

References

- [1] Yang Yong-Jie, Cheng Shao-Jie, Han Guo-Dong 2006 Experimental on acoustic emission during compression rupture procedure of coal sample *Journal of China Coal Society* **31**(5) 562-5
- [2] Zuo Jian-Ping, Pei Jian-Liang, Liu Jiang-Feng 2011 Investigation on acoustic emission behaviour and its time-space evolution mechanism in failure process of coal-rock combined body *Chinese Journal of Rock Mechanics and Engineering* **30**(8) 1564-70
- [3] Chao Ning, Feng Yu, Li-Gang Jing 2011 Experimental research on acoustic emissions characteristic of coal and rock with bursting liability under uniaxial compression *Coal mining Technology* **16**(1) 97-100
- [4] Qin Hu, Huang Gun, Wang Wei-Zhong 2012 Experimental study of acoustic emission characteristics of coal samples with different moisture contents in process of compression deformation and failure *Chinese Journal of Rock Mechanics and Engineering* **31**(6) 1115-20
- [5] Voznesenkii A S, Tavostin M N 2005 Acoustic emission of coal in the post limiting deformation state *Journal of Mining Science* **41**(4) 3-10
- [6] Majewska Z, Mortimer Z 2006 Chaotic behaviour of acoustic emission induced in hard coal by gas sorption-desorption *Acta Geophysica* **54**(1) 50-59
- [7] Cao Shu-Gang, Liu Yan-Bao, Li Yong. 2009 Experimental study on acoustic emission characteristics of coal rock at different confining pressure *Journal of Chongqing University* **32**(11) 1321-27
- [8] Ai Ting, Zhang Ru, Liu Jian-Feng 2011 Space-time evolution rules of acoustic emission locations under triaxial compression *Journal of China Coal Society* **36**(12) 2048-57
- [9] Gao Bao-Bin, Li Hui-Gui, Yu Shui-Jun 2013 Acoustic emission and fractal characteristics of coal rock samples under triaxial compression *Mechanics in Engineering* **35**(6) 49-54
- [10] Su Cheng-Dong, Gao Bao-Bin, Nan Hua 2009 Experimental study on acoustic emission characteristics during deformation and failure processes of coal samples under different stress paths *Chinese Journal of Rock Mechanics and Engineering* **28**(4) 757-66
- [11] Liu Kai-De, Liu Quan-Sheng, Zhu Yuan-An 2013 Experimental study of coal considering directivity effect of bedding plane under Brazilian splitting and uniaxial compression *Chinese Journal of Rock Mechanics and Engineering* **32**(2) 308-16
- [12] Zou Yin-Hui 2003 A study on coal or rock acoustic emission mechanism and relevant experiments *Journal of Xiang Tan Mining Institute* **18**(4) 18-21
- [13] Ji Hong-Guang, Wang Hong-Wei, Cao Shao-Zhong 2012 Experimental research on frequency characteristics of acoustic emission signals under uniaxial compression of granite *Chinese Journal of Rock Mechanics and Engineering* **31**(sup.1) 2900-5
- [14] Tang Shu-Hen, Yan Zhi-Feng, Zhu Bao-Cun 2010 Acoustic emission characteristics of water-saturated coals in uniaxial compression experiment *Journal of China Coal Society* **34**(9) 37-41

Authors	
	<p>Liu Xiaohui, born on October 1, 1977, Sichuan, China</p> <p>Current position, grades: on-the-job Doctor; the lecturer of Xihua University, Sichuan, Chengdu. University studies: Graduated from Sichuan University, Bachelor Degree of Hydraulic Engineering in 2001, Master Degree of Rock Mechanics, mater degree in Geotechnical Engineering from Sichuan University. Scientific interest: geotechnical engineering and water resources and hydropower engineering. Publications: more than 10 papers published Experience: now working in Xihua University, with 8 years teaching experience.</p>
	<p>Zhao Xiaoping, born on May 19, 1984, Jiangxi, China</p> <p>Current position, grades: Chengdu Engineering Corporation, Power China University studies: Bachelor Degree of Engineering at 2009; Sichuan University, Doctor Degree of Engineering (Chengdu, 2013) Scientific interest: rock mechanics Publications: about 10 Experience: Chengdu Institute of Technology University, at Sichuan university since 2009, now working in Chengdu Engineering Corporation.</p>
	<p>Liu Jianfeng, born on August 27, 1979, Henan, China</p> <p>Current position, grades: Sichuan University, Sichuan, China University studies: Graduated from Shandong University of Science and Technology, Bachelor Degree of Engineering in 2002; graduated from Sichuan University, Dr. Degree in Geotechnical Engineering Scientific interest: Geotechnical Engineering, Rock Mechanics and Engineering Publications: about 60 Experience: now working in Sichuan University.</p>

Authors' index					
Azhari Ivan	44	Li Juncheng	79	Wang ZongZhe	118
Cai Sijing	149	Li Shuxiao	86	Wei Xianmin	111
Chang Hongxing	86	Li Wenjing	24	Wu Jinzhao	94
Chen Bingqian	124	Li Xiao-xia	44	Wu Rong-chun	138
Chen Huang	7	Liang Jun	297	Xiang Jinwu	50
Chen Jinbao	263	Liang Tie	283	Xiang Maoqing	36
Chen Jing	44	Liu Chang-liang	19	Xu Fuyuan	230
Chen Ningli	201	Liu D Y	269	Xu Gaochao	73
Chen Shaofei	288	Liu F Y	269	Xu Yueshen	155
Chen Yan	181	Liu Jianfeng	308	Xue C M	174
Cheng Weijun	66	Liu Jun	7	Yang Chenghu	220
Chu Xiao guang	303	Liu Lanying	220	Yang Guojie	283
Chung Peter	191	Liu Lihua	288	Yang Hua	61
Ding Yan	73	Liu Meiling	94	Yang Peng	257
Dong Bin	283	Liu Peng	73	Yang T	174
Dong Yunmeng	73	Liu Qingyou	7	Yang Xianfeng	102
Duan Nannan	230	Liu Shu Jie	55	Yang Xianliang	252
Fang Dianjun	208	Liu Xiaohui	308	Yang Xuefeng	143
Feng Hong	111	Liu Yong	240	Yang Y	174
Feng Zhilin	155	Lu Hong	111	Ye Yanming	155
Fu Xiaodong	73	Luo Jun	149	Yi Haiyan	208
Gong Lixiong	240	Luo Zhangping	50	Yin Wenliang	36
Guo Haixiang	181	Ma Qing	61	Yu Xing	226
Hai Liu	44	Ma Zeng-hui	19	Zeng T T	174
Han Jian-Ning	257	Naibin Jiang	7	Zhang Chaogang	61
He Qian	138	Ni Ming	230	Zhang Feng-li	138
He Xing	7	Nie Hong	263	Zhang Hong	168
Hu Ya Wei	55	Peng Zhaohui	61	Zhang Hong Chao	55
Huang Min	240	Qiu Lirong	131	Zhang Hongyang	29
Huang W	269	Ren Yiru	50	Zhang Jin-bang	138
Jia Lianlian	252	Shen Yiping	86	Zhang Jiong	288
Jia Zhenzhen	275	Shi Guoping	297	Zhang Lei	220
Jiang H F	269	Si Zhaoxia	201	Zhang Lu	257
Jin Song	191	Tang Weidong	94	Zhang Shaoliang	201
Jing Ni-qin	14	Tao Feng	275	Zhang Tianwei	191
Kai Ping-an	19	Wang Cheng	196	Zhang Yingjun	143
Kang Jian-yi	44	Wang Deyun	181	Zhao Jia	73
Kong Xiangsheng	240	Wang Jiangjiang	252	Zhao Xiaoping	308
Kong Ying	303	Wang Jin	162	Zhu Chengfei	86
Li Chao	55	Wang Lijuan	191	Zhu Kejun	181
Li Chun	263	Wang Lin-na	14	Zhu Weidong	162
Li Hong	118	Wang Songling	252	Zhuang Nanjian	50
Li Hong	247	Wang Yan	102	Zou HaiLei	196
Li Jinling	181	Wang Yanhui	149		

Cumulative Index

Mathematical and Computer Modelling

Jun Liu, Xing He, Qingyou Liu, Jiang Naibin, Huang Chen Vibration-modal analysis model for multi-span pipeline with different support conditions*Computer Modelling & New Technologies 2014 18(5) 7-13*

Vibration characteristics analysis is important in the design of multi-span pipeline with different support conditions. In order to analyse the natural frequency and the vibration modal of the multi-span pipeline, a matrix transfer method is proposed in this paper. With the multi-span pipeline divided into single-span pipes, the transmission formulas for the deflection, angle, bending and shear between two adjacent spans are deduced, in combination with the Krylov function solution of the free vibration equation for the single-span pipe, and the constraint condition between the two adjacent spans of the multi-span pipeline. According to the boundary conditions on the starting and ending spans, the natural frequency equation and the vibration modal function between two adjacent spans of the multi-span pipeline are presented. The FORTRAN program based on the above principle is written, and the natural vibration frequencies and the vibration modals of two typical multi-span pipelines are investigated and compared with the results from ABAQUS. It is shown that the model presented in this paper is efficient in the analysis of multi-span pipeline and has the advantages of high computational efficiency and convenience for engineering practice application.

Keywords: Multi-span pipeline, Vibration equation, Natural frequency, Vibration modal

Ni-qin Jing, Lin-na Wang The research of electromotor control based on optimized RBF neural network*Computer Modelling & New Technologies 2014 18(5) 14-18*

RBF neural network suits to control electromotors, which have uncertainty and highly nonlinear systems. However, in practice, RBF neural network also have some obvious defects. For example, the strong dependence on the initial parameter and the poor quality of clustering algorithm. For the above defects, this paper is going to build an optimized RBF neural network through the combination of ant colony optimization algorithms, chaos ergodicity optimization theory and traditional K-means algorithm. On this basis, the optimized RBF neural network will be applied to PID control and then the dynamic performance of the electromotor will be simulationally tested by the designed PID controller. The simulation results show that in the control of electromotor, the optimized RBF neural network has the characteristic of high control accuracy and strong traceability and also it has the ability to guarantee electromotor control system with steady and dynamic performance.

Keywords: RBF neural network, ant colony optimization algorithms, chaos ergodicity optimization, chaos ant colony optimization algorithms, electromotor control

Chang-liang Liu, Zeng-hui Ma, Ping-an Kai A universal tuning method of large dead-time system*Computer Modelling & New Technologies 2014 18(5) 19-23*

The current approach of the PID controller design and tuning for the large dead-time system are almost based on the first-order time delay model, such as all PID algorithms and Ziegler function in MATLAB. And these algorithms generally only apply to the following system $0.1 \leq \tau/T \leq 2$. Therefore, the application effect of these algorithms in large dead-time system are not ideal. Based on the second-order system, this paper proposed a universal tuning method of PID controller for large dead-time process. By the introduction of the controller pre-coefficient K_f , this method makes the large dead-time system PID controller design and tuning simplistic. The fitting formula of controller pre-coefficient K_f was given in this paper. The method is simple, versatile, suitable for the object of $(\tau/T) \rightarrow +\infty$ and second-order, higher-order system, overcomes the limitations of traditional PID control algorithm in large dead-time system applications. The simulation results show that the method is correct, effective and has practical value.

Keywords: dead-time system, PID controller, parameter tuning, second-order system, pre-coefficient

Wenjing Li Using grey-weighted Markov chain model to predict the quantum of highway passenger transport*Computer Modelling & New Technologies 2014 18(5) 24-28*

The grey-weighted Markov model is a prediction model integrating the advantages of grey model and Markov chain model. It can be applied to predict the highway passenger transport quantum. Compared with grey model, the grey-weighted Markov chain model improved the precise of prediction, so the combined model was more appropriate for the prediction of highway passenger transport. Based on the original data of highway passenger transport quantum from 2001 to 2011, the passenger transport quantum in 2012 was predicted with grey-weighted Markov chain model.

Keywords: Grey model, weighted Markov chain, passenger transport quantum, prediction

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(5) 29-35*

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: Modelling in SolidWorks, interface processing, birth-death element, APDL, simulation analysis in ANSYS

Wenliang Yin, Maoqing Xiang Research on the operation modes of hydropower station based on complementary characteristics*Computer Modelling & New Technologies 2014 18(5) 36-43*

Including wind-PV-ES (Wind/Photovoltaic/Energy storage) hybrid power generation system into the scheduling system of grid is the development tendency of safe grid-connection and operation of large wind-PV-ES hybrid power generation system. To solve the active power control problems in hybrid power generation system, this paper analyzes genetic algorithm and quantum genetic algorithm, and also analyzes the importance of energy storing devices in scheduling. Based on this, an optimization model of active power in wind-PV-ES is established. With the expectation of power output fluctuation of the power generation system as the objective function, the optimal scheduling scheme for the model is sought through genetic algorithm and quantum genetic algorithm respectively. The results of Matlab experiment show that the optimal scheduling scheme obtained by means of quantum genetic algorithm is superior to the scheduling scheme obtained by means of traditional genetic algorithm.

Keywords: Hybrid power generation system, Power control, Genetic algorithm, Quantum genetic algorithm

Hai Liu, Jing Chen, Jian-yi Kang, Xiao-xia Li, Ivan Azhari Finite element analysis of the dynamic response of the cardiovascular system to the blunt ballistic impact*Computer Modelling & New Technologies 2014 18(5) 44-49*

On the basis of the Chinese Visible Human Dataset (CVHD), a three-dimensional human finite element model that includes skin, muscle, bone, the lungs, the heart and the vascular trunk was developed. In the LS-DYNA software environment, a numerical simulation of the blunt ballistic impact, which was caused by a 5.56-mm rifle bullet moving with the speed of 910 m/s toward a human torso wearing a composite body armor vest, was performed, and the stress and pressure response of the cardiovascular system were calculated. The simulation results demonstrated that the blunt ballistic impact introduced a high-frequency pressure response on the chambers of heart, which was characterized by a high amplitude and short duration. The peak values of the pressure waves, measured at the ascending aorta and superior vena cava ports, were 659.3 kPa and 542.8 kPa respectively, which suggested that the blunt ballistic impact

on the chest would result in injury to distant target organs through the cardiovascular system. The computational results of this model can provide a basis for predictions of heart injuries, in-depth studies of the mechanical mechanism of cardiovascular injuries to blunt ballistic impacts and further improvements in protective equipment.

Key words: blunt ballistic impact, cardiovascular system, dynamic response, finite element analysis

Nanjian Zhuang, Jinwu Xiang, Zhangping Luo, Yiru Ren Calculation of helicopter maneuverability in forward flight based on energy method

Computer Modelling & New Technologies 2014 18(5) 50-54

A new method for calculating helicopter maneuverability in forward flight is proposed. Empirical equations for evaluating rotor required power are employed. Using energy method, an algorithm to calculate the available overloads, rate of climb and flight trajectory is given. The maneuver performance of AH-1G helicopter is investigated and three kinds of maneuvers including level acceleration, deceleration turning and turning climb followed by accelerating climb are calculated and analysed. Numerical results indicate that the method is effective and feasible, even for three dimensional maneuvering problems. In addition, the method can be applied to predict flight trajectory during forward flight.

Keywords: Helicopter, Maneuverability, Energy Method

Shu Jie Liu, Ya Wei Hu, Chao Li, Hong Chao Zhang Residual life prediction under condition monitoring

Computer Modelling & New Technologies 2014 18(5) 55-60

Reliability assessment and remaining life prediction in the working processes of mechanical products, getting more attention of researchers, can reduce accidents and losses and help improve the preventive maintenance decision-making. This article presents two failure models, linear and exponential, to predict residual life distribution based on the degradation information of mechanical products. Parameters of the models can be estimated using maximum likelihood method. After the real-time monitoring information is acquired, residual life distribution should be updated constantly in order to improve accuracy of the prediction. Experiments were carried out on a double row cylindrical roller bearing to get the vibration information. It proved the validity of the aforementioned method and was applied to compare the two degradation models.

Keywords: reliability, residual life distribution, degradation model, double row cylindrical roller bearing

Qing Ma, Hua Yang, Chaogang Zhang, Zhaohui Peng Effects of global warming for building energy demand in China

Computer Modelling & New Technologies 2014 18(5) 61-65

The impact of global warming on building energy demand in China was investigated by means of whole building energy analysis model and hourly weather data. Four standard multi-story office-building models, representative of four typical climate locations were constructed. For the time period 2050–2100, the climatic temperature scenario models for four typical cities was used that foresees a 2.7-4.2 °C rise in mean annual air temperature relative to the period 1961–1990 normal temperature and is thereby roughly in line with the climate change predictions made by the Intergovernmental Panel on Climate Change (IPCC). The simulation results show that the annual cooling energy demand for office buildings with internal heat gains of 20–30 W/m² will increase by 26-58% while the heating energy demand will fall by 17-52% for the period 2050–2100. This study has also shown that the typical meteorological year (TMY) currently in use by building designers and HVAC engineers in China will lead increasingly to an overestimation of heating energy demand. Similarly, the use of TMY to compute cooling power and cooling energy consumption is likely to result in a progressive underestimation of the future demand. The future building energy demand is set to become a crucial design issue.

Keywords: Global warming, Heating energy, Cooling energy, TMY, Temperature scenarios, Typical office building model

Weijun Cheng On the error rate analysis of dual-hop relaying over composite fading channels using mixture gamma distribution

Computer Modelling & New Technologies 2014 18(5) 66-72

In this paper, we present the end-to-end performance of a dual-hop amplify-and-forward (AF) relaying system over independent non-identical (non-i.i.d) composite Nakagami-lognormal (NL) fading channels by using mixture gamma (MG) distribution. Novel closed-form expressions for the probability density function (PDF) and the moment-generation function (MGF) of the end-to-end signal-to-noise ratio (SNR) are derived. Moreover, the average error rate and the diversity order are found based on the above new expressions, respectively. These expressions are more simple and accuracy than the previous ones obtained by using generalized-K (KG) distribution. Finally, numerical and simulation results are shown to verify the accuracy of the analytical results. These results show that it is more precise to approximate the composite NL distribution by using the MG distribution than using the KG distribution in the performance analysis of cooperative relaying systems.

Keywords: Dual-hop Relaying, Nakagami-lognormal, Mixture Gamma Distribution, Error Rate Analysis

Computer and Information Technologies**Gaochao Xu, Peng Liu, Xiaodong Fu, Yunmeng Dong, Jia Zhao, Yan Ding** A novel task deployment approach based on graph theory for power saving

Computer Modelling & New Technologies 2014 18(5) 73-78

With the increasing of the big datacenter, the power consumption seems to be another overhead except the equipment cost. Saving the power of big datacenter is the hotspot now. In this paper, we proposed TA-BG algorithm based on the linear weighted and graph theory to speed up the execution of tasks. Firstly, utilizing linear weighted to execute first filter to reduce the searching scope for the next research. Secondly, seeking out the hosts that can execute tasks fast based on graph theory. Finally, placing the host on the hosts selected above. The experiments indicate that TA-BG can save power of datacenter by reducing the executing time. Besides, the TA-BG even performs well on load balance.

Keywords: Cloud Data Centre, Task Allocation, Power Saving, Graph Theory

Juncheng Li Image enlargement based on the hyperbolic Coons interpolation

Computer Modelling & New Technologies 2014 18(5) 79-85

A method for image enlargement, making use of the hyperbolic Coons interpolation surface with shape parameters, is investigated in this paper. As a non-polynomial model, the hyperbolic Coons interpolation surface can represent the image better than the general interpolation methods. By altering the values of the shape parameters, the effects of image enlargement can be adjusted until achieving the satisfactory results. Experimental results show that the effects of image enlargement making use of the hyperbolic Coons interpolation surface are better than the general interpolation methods.

Keywords: image enlargement, hyperbolic Coons interpolation surface, shape parameters

Yiping Shen, Shuxiao Li, Chengfei Zhu, Hongxing Chang A fast top-down visual attention method to accelerate template matching

Computer Modelling & New Technologies 2014 18(5) 86-93

This paper presents a fast top-down visual attention method to downsize the search space of template matching. Such a method first generates patterns representing the local structures, and then calculates the pattern distributions representing the template and its surroundings. From here two separate operations are performed: the "pattern weight" is first introduced, which describes how well a certain pattern is correlated to the template, and then weights of all patterns are calculated for later reference. This is the "off-line" operation, and in comparison the "on-line" operation only calculates the pattern of each pixel, whose weights can be indexed conveniently from the off-line results. With all pixels' pattern weights calculated, the weight image is ready, from which we can extract the region of interest for subsequent matching. Experiments showed that our method obtained at least 6.21 times speed-ups over the state-of-the-art methods with little or no loss in performance.

Keywords: template matching, visual attention, top-down attention, saliency, region of interest

Weidong Tang, Jinzhao Wu, Meiling Liu Step semantics and action refinement in event structures*Computer Modelling & New Technologies 2014 18(5) 94-101*

An event structure acts as a denotational semantic model of concurrent systems. Action refinement is an essential operation in the design of concurrent systems. However, there exists an important problem about preserving equivalence under action refinement. If two processes are equivalent with each other, we hope that they still can preserve equivalence after action refinement. In linear time equivalence and branching time equivalence spectrum, step equivalences, which include step trace equivalence and step bisimulation equivalence are not preserved under action refinement [17]. In this paper, we define a class of concurrent processes with specific properties and put forward the concept of clustered action transition, which ensures that step equivalences are able to preserve under action refinement.

Keywords: event structure, action refinement, concurrency, step equivalence, clustered equivalence

Xianfeng Yang, Yan Wang A research into static traffic routing and resource optimization algorithm based on genetic and tabu search*Computer Modelling & New Technologies 2014 18(5) 102-110*

In order to solve the issue of optical network's static traffic routing and resource optimization, this paper puts forward a hybrid genetic and tabu search virtual reconfiguration algorithm (HGTS-VRA) and designs the key elements. This algorithm could effectively integrate the large scale searching ability of genetic algorithm and the outstanding local searching ability of tabu search algorithm. The simulation comparison result and analysis result show that the HGTS-VRA put forward by this paper enjoys excellent advantages in the field of traffic routing and resource optimizing. In addition, it offers outstanding extendibility and robustness.

Keywords: Hybrid Genetic, Tabu Search, Static Traffic Routing, Resource Optimization

Xianmin Wei, Hong Lu, Hong Feng ABMP: adaptive bitmap protocol within TDMA for mobile underwater sensor networks*Computer Modelling & New Technologies 2014 18(5) 111-117*

Media access control (MAC) protocol design is one of hot topics in the research of underwater acoustic sensor networks (UWSN). The major challenge is the phenomenon of space-time uncertainty caused by the long delay in underwater signal propagation, where the occurrence of frame conflict is determined by not only two nodes' transmission time, but also by their locations. In this paper, the adaptive bitmap protocol-based (ABMP) within time division multiple access (TDMA) was proposed for UWSN, with specialized space-time uncertainty among mobile underwater nodes reducing channel idle time and improving energy efficiency and transmission efficiency. Finally, simulation experiments are conducted to present that the proposed protocol has better communication efficiency and energy efficiency, compared with other MAC protocols of Token-TDMA and T-lohi in terms of network traffic, end-to-end delay and energy efficiency.

Keywords: Underwater sensor networks, Adaptive bitmap protocol, Media access control, Time division multiple access

Hong Li, ZongZhe Wang Design of grey PID controller for DC Servo Motor*Computer Modelling & New Technologies 2014 18(5) 118-123*

Concerning the uncertainties may be existed in DC motor servo system and the control quality that the external disturbance may influence the control algorithm of traditional PID, a kind of PID control algorithm based on grey prediction theory was proposed. With the grey theory's processing ability on the unknown information data, the algorithm established the grey model for the uncertainties, and real-time compensated the system's unmodelled feature and disturbance signal, thus improving the control precision. The simulation result can show that the proposed PID control algorithm based on grey prediction theory can effectively predict and compensate the unmodelled feature and disturbance signal in DC motor servo system and improve the control precision of the controller, thus providing the theoretical basis for the industrial application of PID control algorithm based on grey prediction theory.

Keywords: PID control, DC motor, Noise, Disturbance, Grey prediction

Bingqian Chen A new automatic registration method for InSAR image based on multi-step strategy*Computer Modelling & New Technologies 2014 18(5) 124-130*

Interferometric Synthetic Aperture Radar (InSAR) technology has been widely used in various applications. The registration of SAR (Synthetic Aperture Radar) images is the first step in interferometric processing therefore accurate registration is essential for the successful creation and interpretation of interferometric products. However, with the growing number of SAR satellite launch and the amount of data acquisition, the degree of automation of image registration have become increasingly demanding. In this paper, we propose an automatic registration approach based on multi-step matching strategy. In the first step, key points are detected and matched using modified scale invariant feature transform (SIFT) operator which modified by us reducing the influence of speckles. In this step, owing to the existing of speckle and the defect of matching strategy of SIFT operator, the expected level of matching accuracy is about 2 to 3 pixels. In the second step, correlation matching (CM) is used to exclude the matched points with low correlation. In the third step, the probability relaxation (PR) algorithm based on global matching is used to induce consistency constraint and ensure reliability of the matching result. Finally, corresponding transformation function is determined through the relationship established by matched point pairs. In order to verify the applicability of proposed methodology, two SAR images acquired over mountainous regions are used in our experiment. The experiment results show that subpixel registration accuracy and good efficiency have been achieved, which demonstrates the correctness and feasibility of proposed method.

Keywords: InSAR, Image registration, Feature detection, SIFT, Correlation matching

Lirong Qiu An approach for reference model implementation by predicting all possible output of design*Computer Modelling & New Technologies 2014 18(5) 131-137*

In verification system, it is preferable to build reference model at transaction level which does not produce the output as the same latency as the design itself. But due to the lack of accurately modelling design's behaviour, there are some scenarios that design's output is different with reference model's output due to the different processing delay of stimulus. Scoreboard can get lots of comparison failure when it tries to do comparison between the output of reference model and design under such scenarios. In this case, neither reference nor design is wrong from functionality, but output comparison failure will mix up with the true design issue and bring trouble to the automatic check on design's behaviour. Cycle based reference model does not have such problem. But it usually takes great effort to implement cycle based reference models and maintain them. This paper provides its study on implementation style of reference model. By predicting all possible output of design, this paper presents a method for reference model to handle such stimulus competition scenarios at the transaction level. The paper also discusses the reference model's reaction effect on generator, which helps the test hit design's corner case.

Keywords: System Verilog, reference model, scoreboard, competition stimulus, coverage driven verification

Rong-chun Wu, Feng-li Zhang, Jin-bang Zhang, Qian He Application of fuzzy comprehensive evaluation in weapon equipment systems*Computer Modelling & New Technologies 2014 18(5) 138-142*

Analysis and evaluation of the operational effectiveness of weapon equipment operational systems has always been a complex problem, a study of its evaluation technology is of great significance. Task oriented operation, this paper discusses equipment operational system dynamic integration needs, and discusses the steps and comprehensive performance evaluation method of weapon equipment operational system, made a fuzzy comprehensive evaluation method to better adapt to comprehensive performance evaluation of weapon equipment operational system. According to the characteristics of system evaluation factors, and gives an indicator of quantitative methods established based on analytic hierarchy process and correlation analysis of comprehensive performance evaluation model, based on three types of weapon equipment operational system data, for example, proves the validity of the method.

Keywords: Weapon Equipment Systems, Comprehensive Performance, Combat Effectiveness, Multi-level Fuzzy comprehensive evaluation

Operation research and decision making**Yingjun Zhang, Xuefeng Yang** Expert system based on fuzzy rules for maritime search and rescue*Computer Modelling & New Technologies 2014 18(5) 143-148*

Search and rescue (SAR) plan decides the result of SAR activity, and relates with the safety of life and property at sea. To improve the efficiency and standard of SAR, and give the SAR officer support to make better decisions, expert system (ES) is been researched by this paper, and the ES based on fuzzy rules for maritime SAR is proposed. Firstly, the structure of ES based on fuzzy rules for SAR is designed. Secondly, we have researched SAR knowledge acquisition and knowledge representation, chosen five ways to acquire SAR knowledge. At last, we designed the inference engine of ES for SAR, and introduced it in an example.

Keywords: fuzzy rules, expert system, supporting system, search and rescue

Jun Luo, Sijing Cai, Yanhui Wang Research on dynamic risk identification model of shield tunnelling based on REASON model*Computer Modelling & New Technologies 2014 18(5) 149-154*

Aiming at the dynamic risk identification problem in shield tunnelling, and with the lack of research on dynamic risk identification theory and human factors in shield tunnelling, an analysis model of shield tunnelling based on REASON model has been proposed to establish in this paper. Relying on the fault tree theory and the model that established, the accident rule base has been built. After forming the REASON model into a network, the dynamic risk identification model for shield tunnelling has been built to provide theoretical guidance for dynamic risk management during the construction.

Keywords: Shield tunnelling, REASON model, dynamic risk identification model, risk management

Yanming Ye, Yueshen Xu, Zhilin Feng Tag-based process recommendation for social business process modelling*Computer Modelling & New Technologies 2014 18(5) 155-161*

Social BPM (Business Process Management) has become a new research hotspot in business process management field because of its capability of handling the flexibility and dynamics of process in social circumstance by means of integration of social software and BPM. The key technique is process modelling, and note worthily the process modelling is more complex in social BPM than in traditional BPM. This paper presents the definition of social business process model to identify the difference with the traditional business process model and gives a prototype of social BPM system. The modeller in the prototype system may be a common user without professional knowledge, so a tag-based process recommendation method is proposed to facilitate modelling. The experiment result shows that the method is valid and effective in computer-aided intelligent process modelling.

Keywords: process recommendation, process modelling, social business process model

Jin Wang, Weidong Zhu The impact of capital structure on corporate performance based on panel threshold model*Computer Modelling & New Technologies 2014 18(5) 162-167*

This paper takes Value Added as an indicator of corporate performance. In considering the case of differences in growth opportunities, we use Panel Threshold Model to do an empirical analysis in the relationship between environmental management, capital structure and corporate performance of listed companies in China. The results show that: There is a weak positive correlation between environmental management and corporate performance, which means environmental management can improve corporate performance; there is a significant regime effects between capital structure and corporate performance relying on corporate growth opportunities. Capital structure and corporate performance are negatively related for low growth companies and positively related for high growth companies. This is consistent with the classical theory of capital structure, which means Value Added is more suitable as an indicator of corporate performance than profits in China.

Keywords: Panel Threshold Model, Value Added, Environmental Management, Capital Structure, Growth Opportunities

Hong Zhang Short-term prediction of wind power based on self-adaptive niche particle swarm optimization*Computer Modelling & New Technologies 2014 18(5) 168-173*

Connecting wind power to the power grid has recently become more common. To better manage and use wind power, its strength must be predicted precisely, which is of great safety and economic significance. Speed sensors are widely applied, it make prediction of wind power more accurate. In this paper, the short-term prediction of wind power is based on self-adaptive niche particle swarm optimization (NPSO) in a neural net. Improved PSO adopts the rules of classification and elimination of a niche using a self-adaptive nonlinear mutation operator. Compared with the traditional method of maximum gradient, NPSO can skip a local optimal solution and approach the global optimal solution more easily in practice. Compared with the basic PSO, the number of iterations is reduced when the global optimal solution is obtained. The method proposed in this paper is experimentally shown to be capable of efficient prediction and useful for short-term power prediction.

Keywords: Speed sensor, PSO, Niche, Short-term power prediction, Neural net

Chengmeng Xue, Yu Yang, Tao Yang, Tingting Zeng Person-organization fit evaluation and process optimization based on the matching theory*Computer Modelling & New Technologies 2014 18(5) 174-180*

To achieve an optimal bidirectional person-organization fit (P-O fit) and improve the overall satisfaction degrees for both of the two sides, a bidirectional P-O fit evaluation and process optimization model is established based on the Matching Theory in this paper. To begin with, the bidirectional P-O fit evaluation factors set is built after the analysis of the indexes of these factors, and the index weights are calculated with the Rough Set Theory; Then, a Bidirectional P-O fit Evaluation and Process Optimization Model is proposed, with the Fit Conflict Resolve Algorithm (CRA) to ensure the persons and organizations to be matched one-to-one; Finally, the validity of this model is verified by its implementation in the enterprise HXMS.

Keywords: persons and organizations, bidirectional fit, fit degree, Cauchy Distribution Function

Jinling Li, Haixiang Guo, Yan Chen, Deyun Wang, Kejun Zhu An artificial fish swarm algorithm for solving a bi-objective capacitated vehicle routing problem*Computer Modelling & New Technologies 2014 18(5) 181-190*

The paper focuses on a capacitated vehicle routing problem with two objectives: one is attainment of specific load factor and the other is minimization of total travel cost. Our approach is based on artificial fish swarm algorithm, a swarm-based heuristic, which mimics the foraging behaviour of a fish swarm. After initializing a school of artificial fish, whose validity is guaranteed by a designed repair operator, global optimal solution search is processed through random behaviour, prey behaviour, swarm behaviour, and follow behaviour. Experimental results for a practical distribution instance are reported and show that the artificial fish swarm algorithm performs better than sweep algorithm and genetic algorithm. This paper contributes to the solution methods of vehicle routing problem.

Keywords: Vehicle routing problem, Artificial fish swarm algorithm, Sweep algorithm, Genetic algorithm

Lijuan Wang, Song Jin, Tianwei Zhang, Peter Chung Analysis of price rate models for household water consumption in urban China*Computer Modelling & New Technologies 2014 18(5) 191-195*

Price rate models are used for a variety of purposes including water conservation, what's more, it is an effective and feasible method to establish a water-conserving society. The survey of household water consumption was conducted in a residential community in Hebei Province. An analysis of the survey data shows that an average family's water consumption is 6.4t per month, the standard is level 3. Three price rate models are proposed, the principle of the first model is the decrease of water consumption in level 1 and no increase in level 2, the second model is no change in level 1 and an increase in level 2, the third model is no increase for 80% of families and only 20% increased. After an analysis and comparison of water consumption and expenses corresponding to the three models, this paper presents

these three models as a valuable reference for department decisions. The results of the first model should be adopted in Hebei Province, the demarcation point of every level is 8t and 13t, the price of every level is 3.04 RMB/t, 4.56 RMB/t and 9.06 RMB/t.

Keywords: Price rate model, Resident water consumption, Data survey, Water conservation

HaiLei Zou, Cheng Wang Calculation of China's environmental efficiency based on the SBM model with undesirable outputs

Computer Modelling & New Technologies 2014 18(5) 196-200

With the rapid development of china's economy, the environment of china faces some prominent questions, industrial pollution, water pollution, serious smog and other problems continue get worse, therefore, it's very significant to analyse China's regional environmental efficiency. This paper mainly measure china's regional environmental efficiency by SBM model with undesirable outputs, and according to the calculation results, we find that the overall average level of china's environment efficiency is low, and the gap between different provinces is large. Finally, this article gives some policy proposal about how to increase china's environmental efficiency and reduce pollution emission.

Keywords: Environmental Efficiency, SBM Model, Undesirable Outputs

Zhaoxia Si, Shaoliang Zhang, Ningli Chen Correlation model analysis on the land price fluctuations in Beijing and Tianjin City in China

Computer Modelling & New Technologies 2014 18(5) 201-207

According to the indexes of the city land price, the co-integration analysis, the Granger causality test, the impulse response function and the variance decomposition method were used in this paper to analyse the correlation model of the land price fluctuations in Beijing and Tianjin city in China. The results showed that the land prices in Beijing and Tianjin city had a long-term co-integration relationship. The Granger test model showed that the land prices in Beijing and Tianjin city was in line with positive correlation bidirectional causality. In addition, the raise of 1% land price in Beijing city caused a raise of 0.96% in Tianjin city. Conversely, a raise of 1% in Tianjin city caused a raise of 1.03% in Beijing city. By comparing the mutual influence degree, the land price fluctuations in Beijing city had a greater influence on that in Tianjin city.

Keywords: Land price fluctuation model, Co-integration test, Granger causality model, Impulse response function

Haiyan Yi, Dianjun Fang Dynamic evaluation and simulation of variant-driven complexity costs in multi-echelon automotive supply chains

Computer Modelling & New Technologies 2014 18(5) 208-219

The proliferation of product variety driven by part variants imposes a great impact on the costs, performance and ecological burden of automotive supply chains. Costs are always the first aspect to be considered by manufacturers. This paper creates a multi-echelon automotive supply chain in the light of the process model and improves the KPI of low costs to evaluate the variant-driven complexity costs. Then, the evaluation models of different complexity costs are constructed. An automotive supply chain scenario model is constructed with the OTD-NET and the simulation results are analysed. The study of this paper will be helpful to automotive manufacturing in seeking to optimize the number and range of product variants.

Keywords: Automotive Supply Chain, Product Variety, Complexity Costs, OTD-NET

Chenghu Yang, Lanying Liu, Lei Zhang Optimal acquisition and pricing policies for remanufacturing systems with initial investment

Computer Modelling & New Technologies 2014 18(4) 220-225

The problem of used product (core) acquisition is an important issue in remanufacturing. In traditional models, remanufacturing systems are assumed to be well-established. However, remanufacturing systems in most developing countries are imperfect and remanufacturers need to make huge initial investments to improve the remanufacturing

systems. It is therefore suggested that the effects of initial investment cannot be neglected. In this paper, an acquisition and pricing problem in imperfect remanufacturing systems is studied. The problem is firstly formulated as a two-period nonlinear programming model, and the closed forms of the optimal solution are presented based on Karush-Kuhn-Tucker conditions. Next, the multi-period acquisition and pricing problem, and the effects of the initial investment are discussed. Finally, the conclusions are testified by numerical examples. The results show that with the remanufacturing system improved, the remanufacturer will increase investment and acquire more cores.

Keywords: Core acquisition, Imperfect remanufacturing systems, Nonlinear programming model, Karush-Kuhn-Tucker conditions

Xing Yu Continuous-time optimal portfolio model with mean-reverting process

Computer Modelling & New Technologies 2014 18(5) 226-229

This paper studies a continuous-time portfolio optimization problem. It is proposed a simple but powerful approximation approach that is both accurate and computationally efficient for the terminal expectation of the investors with mean-reverting process, which is different from the existing literatures that apply the dynamic programming method. Numerical examples illustrate the computational efficiency and accuracy of our approach when compared with results from Monte Carlo (MC) simulations.

Keywords: Continuous-time portfolio, Mean-reverting process, Optimization; Monte Carlo

Nannan Duan, Fuyuan Xu, Ming Ni Evolutionary game analysis of enterprises' technological innovation strategies

Computer Modelling & New Technologies 18(6) 230-239

Different from general studies on competition and/or cooperation relationship of enterprises, the paper classified enterprises' technological innovation strategies into cooperation, neutrality and competition, and analysed the evolution of enterprises' relationship and strategy selection in technological innovation using the evolutionary game method and the MATLAB simulation technique. The paper drew the following conclusions: (1) the increase in technological content may cause two uncertain cases of enterprises' strategy, namely both parties chose the cooperation strategy, or one party chose the cooperation strategy while the other party chose the noncooperation strategy; (2) the increase in innovation revenue (coefficient) prompted enterprises to eventually tend towards the cooperative innovation strategy, no matter what the initial relationship between enterprises was; (3) the increase in the number of enterprises with the neutrality attitude promoted enterprise groups to tend towards cooperative innovation.

Keywords: technological innovation, evolutionary game, cooperation, neutrality, competition

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

Lixiong Gong, Xiangsheng Kong, Yong Liu, Min Huang Subpixel edge extraction of part ant colony optimization-based and dimensional measurement

Computer Modelling & New Technologies 2014 18(5) 240-246

Put forward a method combined improved ant colony and Zernike moment to detect image subpixel edge aiming at traditional ant colony algorithm's drawback of long time consumption and easily to be affected by noise. The methods improved parameters from clustering centre setting, clustering operator selecting and pheromone updating, then extracted subpixel image edge based on Zernike moments. Therefore, the result of image edge extraction is good and effective. Lastly, least square fitting is used to locate coordination of image edge and bearing of SKF 32308 J2/Q dimensions such as inner and outer diameters were measured. The result shows that the algorithm proposed can well to measure circular parts dimensions and has high precision.

Keywords: Subpixel, Ant Colony Optimization, Bearing, edge detection

Hong Li Application of PID-type iterative learning control for DC motor

Computer Modelling & New Technologies 2014 18(5) 247-251

Iterative learning control is a new control technology, which is a branch of intelligent control theory and particularly suitable for the controlled object with repetitive motion. In this paper, a PID-type iterative learning control for DC

motor based on the characteristics of repeating motion of DC motor was proposed and the convergence of iterative learning control algorithm was analyzed. The input of controlled system in current cycle was amended by the error achieved between the system output and the desired trajectory in previous iteration. It was tested that PID-type ILC had good performance and stability through a large number of simulations and the experiments of the velocity tracking are done by MATLAB software. The results showed that the velocity tracking precision of DC motor was higher and the error was smaller with the increasing number of iterations. The velocity tracking error was close to zero. It was also shown that the motor could fully track the given desired trajectory in some certain iteration. It was also revealed from simulation results that the proposed control strategy was valid and effective for the DC motor.

Keywords: PID control, DC motor, PID-type iterative learning control

Xianliang Yang, Lianlian Jia, Songling Wang, Jiangjiang Wang Based on pressure gradient model to determine leakage point in heating pipe network

Computer Modelling & New Technologies 2014 18(5) 252-256

This paper describes a calculation method based on pressure gradient model to determine leakage point in central heating pipe network which decreases the effect of the ratio frictional resistance. In the calculation method, a pipe resistance characteristic coefficient is introduced. This characteristic coefficient is corrected by hybrid adaptive genetic algorithm. With the characteristic coefficient and the pressure value of each node in heating pipe network, the leakage point orientation in central heating pipe network can be done with the positioning analysis on the heating pipe network leak. The pressure value is calculated using the theory of graph algorithms. In this analysis, the pressure gradient is the most important analytical method. According to the data monitored by supervisory control and data acquisition (SCADA) system, the leak position can be located. The experimented results show that the method to locate the leakage point in the thermal system pipe network meets the theoretical need and the experimental one.

Keywords: Heating pipe network, Leakage location, Drag coefficient, Pressure gradient

Jian-Ning Han, Peng Yang, Lu Zhang External locating of moving targets for 3D IMRT using parallax method

Computer Modelling & New Technologies 2014 18(5) 257-262

Due to the complexity of intensity modulated radiotherapy, it is difficult to meet the current situation of treatments which require real time locating of moving target. In this paper, we propose a new method based on parallax method using external label to locate the moving target. A column test-piece is used to simulate the human body structure for data analysis. A defect model for human body target is implanted inside the test-piece, which is the moving target. Based on the parallax method, the depth of the defect can be obtained by using the two images captured before and after the test-piece moves a short distance. The possible errors which affect the test results are analysed. The effects of errors can be reduced by adjusting the system parameters. The results show that the parallax method is a simple but efficient approach, which can be used for locating moving target in intensity modulated radiotherapy.

Keywords: Intensity modulated radiotherapy, Moving target location, Parallax method, Three-dimensional location, External Label

Chun Li, Hong Nie, Jinbao Chen Motion analysis and simulation of a 12-Tetrahedral Walker Robot

Computer Modelling & New Technologies 2014 18(5) 263-268

A novel robot mechanism-tetrahedral rolling robot is introduced in the paper. The robot comprises of 26 extension struts and 9 nodes. When the COG of tetrahedron exceeds the stability region, the robot will roll. The structure of the 12-TET robot is described. Designing method of the robot is given, and it is proved correct and feasible through simulation. Kinematic models in different motion phases are analysed in the paper, and the rolling critical condition is formulated. The effectiveness of the method is testified through simulation. The study of the paper will provide important reference for the dynamic analysis, optimization design and control of the tetrahedral rolling robot.

Keywords: Variable tetrahedron robot, Gait planning, Motion analysis

W Huang, D Y Liu, H F Jiang, F Y Liu Numerical and analytical solution of stresses on a box-type lining structure under the effect of ground fracture

Computer Modelling & New Technologies 2014 18(5) 269-274

This study attempts to study the stress mechanism of a box-type lining structure during its inclined penetration of ground fissure and calculate the normal stress and shearing stress of the structure section. Based on thin-wall structure theory combined with the stress boundary conditions of the physical model of the box-type lining structure, we derived the analytical solution of the normal stress and shearing stress of the physical model. The stress analytical solution indicates that the damage of the footwall of the ground fissure is more serious than that of the hanging wall, which could match the physical model experiment. The effectiveness and accuracy of the analytical solution of the normal stress and shearing stress of the section were verified using finite element software to establish the mechanical model of the box-type lining structure. The results of the numerical model were compared with the analytical solution results.

Keywords: Box-type structure, Lining structure, Ground fissure, Analytical solution

Zhenzhen Jia, Feng Tao Numerical simulation on methane explosion propagation in a one-dimensional straight duct with porous metal materials

Computer Modelling & New Technologies 2014 18(4) 275-282

Based on the theoretical and experimental results of methane explosion propagation in porous metal materials, methane explosion propagation in a one-dimensional straight duct with different layers of porous metal materials is simulated by *Fluent* software. The layers and length of porous metal materials have influence on the flame propagation velocity and explosion shockwave overpressure. Only the propagation distance of methane explosion is beyond 5 times the diameter of the duct, the flame and explosion wave can be attenuated by the porous metal materials. Moreover, the more the layers are, the better the attenuation effect is. The numerical simulation results show that, during methane explosion propagation in porous metal materials, the attenuations of explosion wave overpressure and flame propagation velocity takes on synchronization and correspondence. Consequently, the porous metal materials can suppress methane explosion propagation. The process of methane explosion in the duct well reappears with numerical simulation, thus the model is established correctly and the numerical simulation is a good supplementary means of experiment.

Keywords: Methane Explosion, Porous Metal Materials, Flame Propagation Velocity, Explosion Shockwave Overpressure, Fluent Software

Bin Dong, Guojie Yang, Tie Liang Research and realization of handheld radio direction finding communication system

Computer Modelling & New Technologies 2014 18(5) 283-287

Radio direction finding communication system has been widely used in civil and military fields. The handheld communication equipment with small development volume and low power has become the irresistible trend. Combined with magnitude comparison, a handheld direction finding and communication receiving system is researched and designed. And it is applied to radio communication system. This system can remove the influence of antenna on the hardware circuit, human body interference and surrounding environment. Besides, it has the advantages of small volume, fast direction finding speed and good capacity of resisting disturbance. Therefore, this handheld radio direction finding communication system has significant application value in public security, forest fire prevention and geological prospecting etc.

Keywords: radio, direction finding, communication, magnitude comparison

Jiong Zhang, Shaofei Chen, Lihua Liu Research on commutation torque ripple suppression strategy of BLDCM based on iterative learning

Computer Modelling & New Technologies 18(5) 288-296

Brushless DC motor (BLDCM) is the DC motor which uses electronic commutation instead of mechanical commutation. The torque ripple caused by mechanical structure of the motor is small, while the torque ripple during the commutation is about 50% of the average torque. By analysing the unique features of commutation torque ripple, we can obtain the relationship between the commutation current and torque ripple. The current ripple can be suppressed through adding voltage compensation of the three-phase bridge inverter circuit. Then the commutation

torque ripple suppression strategy based on iterative learning was raised in this paper. With the MATLAB Simulink platform, BLDCM simulation models, which based on the iterative learning are established and simulation contrast experiments of open or closed-loop iterative P-type with or without adding the voltage compensation have been designed to test and verify the effectiveness of the suppression strategy.

Keywords: BLDCM, Iterative learning, Torque Ripple, Voltage compensation

Guoping Shi, Jun Liang The gradual learning static load modelling method based on real-time fault recorder data

Computer Modelling & New Technologies 18(5) 297-302

Setting a real-time load model is an effective way to overcome time-variation of power load in course of power load modelling. On the basis of load data sorting, this paper proposes a gradual learning static load modelling method based on power fault recorder data. Firstly, power fault recorder collects and stores valid load data. Secondly, all valid load data will be classified by the time, static load model can be built corresponds to each classification. Thirdly, model parameters of each sort are identified by gradual learning method, for the goal of global fitting optimal for the measured active power and calculated active power, the load model parameters are optimized by using curve fitting method. The identified model parameters can be applied to power system calculation directly without preserving all load data, essential feature of all load data is reserved and modelling operational efficiency is improved greatly. Simulation results show that the gradual learning method is right and effective, which is easier to realize and is of higher precision compared with least squares method, therefore the method has widely applicable value and is prospective in power system on-line static load modelling.

Keywords: Fault Recorder, Static Load Modelling, Parameter Identify, Gradual Learning, Curve Fitting

Ying Kong, Xiao guang Chu The scroll flow and torque prediction with the wavelet neural network optimized by PSO and BP

Computer Modelling & New Technologies 2014 18(5) 303-307

A new Compressed Air Energy Storage(CAES) with scroll was proposed to promote the storage efficiency, which can be acquired by the scroll efficiency tracking control with the timely evaluation, but the flow and torque is not easy acquired because the sensors possessed the merits of high price, lower life-span and subjection to the disturbance, so a torque and flow prediction algorithm based on Wavelet Neural Network (WNN) is proposed adopting a hybrid learning algorithm combining Particle Swarm Optimization (PSO) with BP. Through the comparison between predictive and the experimental data and the scroll efficiency experiment, the proposed prediction method is validated and can be successfully used to improve Pneumatic conversion efficiency.

Keywords: Compressed Air Energy Storage, Scroll, Particle Swarm Optimization (PSO), Wavelet Neural Network

Xiaohui Liu, Xiaoping Zhao, Jianfeng Liu A study on the acoustic emission characteristics of the coal rock on different bedding direction

Computer Modelling & New Technologies 2014 18(5) 308-313

The rock mechanics servo system (MTS815) and the acoustic emission testing system (PCI-2) were selected in this paper to conduct the uniaxial compression test and the acoustic emission test from two different direction of the coal and to clarify the acoustic emission feature during the process compression deformation of the coal. The results show that there is an obvious difference on the characteristic parameter of acoustic emission and the spatial distribution during the process of uniaxial compression, which is from different directions. On the initial compression phase, the stress growth rate of the coal from the vertical direction is greater than that from parallel bedding coal and the acoustic parameters. Acoustic emission ringing, energy and event count rate are less than the coal rocks from parallel bedding. The acoustic emission count will increase with arise stress, the coal from the parallel direction is more stable than the coal from the vertical direction; the coal acoustic emission has significantly reduced which from the vertical direction when after the peak stress and there is a little change with the parallel coal rock; The acoustic emission event of the coal which is from the vertical direction and the parallel coal usually concentrates in the lower part of the coal.

Keywords: Coal Rock, Acoustic Emission, Spatial Distribution of Events, Bedding Direction, AE Characteristic Parameters