



COMPUTER MODELLING
AND
NEW TECHNOLOGIES



2014
VOLUME 18 NO 2

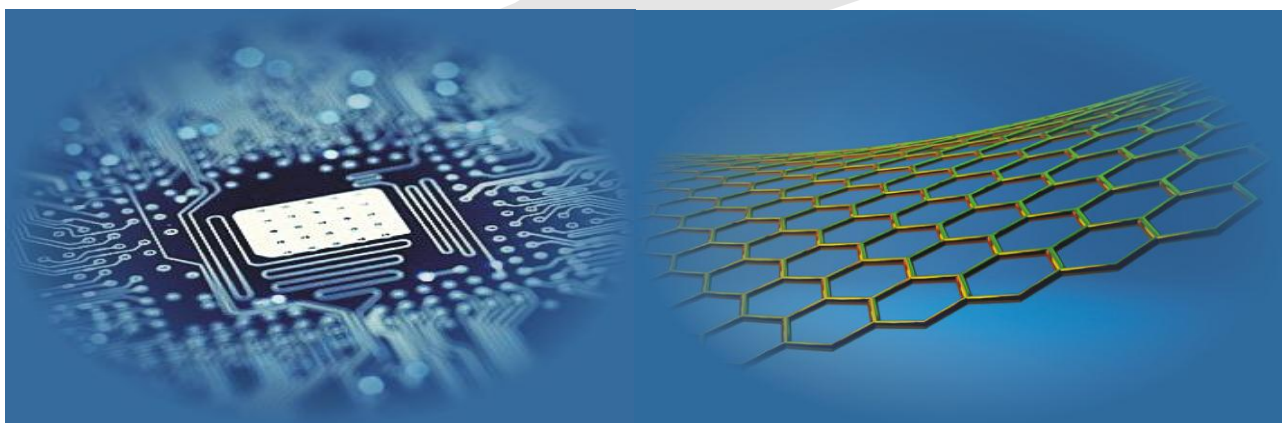
ISSN 1407-5806 ISSN 1407-5814 on-line

Transport and Telecommunication Institute
and
Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2014 Volume 18 No 2

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. Eugene Kopytov	<i>Transport & Telecommunication Institute, Latvia</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>
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 innovative technologies operation research and decision making nanoscience and nanotechnologies operation research and decision making 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.2 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) 67100593 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/ Ex Compindex



Content

Editors' Remarks		5
Nanoscience and Nanotechnology		
Yu N Shunin, S Bellucci, Yu F Zhukovskii, V I Gopejenko, N Burlutskaya, F Micciulla, T Lobanova-Shunina, A Capobianchi	CNT-Fe-Pt interconnect electromagnetic simulations for magnetically stimulated CNT growth and novel memory nanodevices	7
Fundamental Nature Phenomena		
A Ershkovich	On the classical roots of the Schroedinger equation	24
Mathematical and Computer Modelling		
Xiao Liang, Wei Li, Weitong Fan, Guocheng Zhao	Numerical simulation and hydrodynamic performance prediction for hydroplane longitudinal motion	27
Song Xiao-ru, Chen Hua, Song Bao-wei, Wu Jie	Research on adaptive AUV tracking control system based on least squares support vector machine	33
Yaping Zhang, Chunxiao Liu, Hong Zhang, Guozhu Cheng	A study on capacity of signalized intersection under snow-ice conditions based on classical model modification	40
Zhang Hong, Li Jie, Lv Yuejing	Structure performance analysis of vehicular <i>ad hoc</i> networks based on the complex network theory	46
Xingping Wei, Jun Li, Xiaoyu Liu, Shiping Zhao	A fast on-line two-dimensional sizes measurement method for micro part	52
X-Y Meng, Y Liu, Y-X Gao, M Wang, Z-H Liu	Basic set of equations in mesoscale meteorological model associated with local fractional derivative operators involving the Cantorian and Cantor-Type spherical coordinates	57
Junsheng Jiao	DSM control strategy of solid oxide fuel cell distributed generation system	63
Jing Guoqing, Liu Guixian, Xu Yang, Zhang Jiong	Microanalysis of sub-ballast direct shearing under normal stress	70
Liu Xiaoyong, Su Tiexiong, Zhang Yi	Numerical analysis and improvement of torsional vibration of shaft systems for engine with cylinder deactivation	76
Gong Yusheng	Simulation design of Bursa-Wolf coordinate transformation model based on the access	83
Nahla Mohamed Ahmed, Ling Chen	New approaches for link prediction in temporal social networks	87
Liu Ranran, Li Zhengming	Modelling study of p-xylene oxidative side-reaction based on Bayesian filtering	95
Huang Li	Bipartite graph model for RDF data cleansing	99
Information and Computer Technologies		
Yang Xianhui, Ren Hongge, Jing Weipeng	A dynamic MAC protocol for wireless sensor network based network traffic monitoring and feedback	107
Guoyan Yang, Xin Guan	A concurrent MAC protocol based on location information in wireless sensor networks	114
Huisong Li	Flow-insensitive type qualifier inference on programming languages allowing type casts	120
Lichuan Gu, Jianxiao Liu, Chengji Wang	Fragments based tracking with adaptive multi-cue integration	126
Shaobo Shi, Qi Yue, Qin Wang	FPGA based accelerator for parallel DBSCAN algorithm	135
Kanghong Duan, Hongxin Zhang, Shilin Song, Peigang Wang	Micro real-time pre-emption operating system for industry wireless sensor networks	143
Ao Jianfeng	Ground crack extraction in mining subsidence areas based on point cloud data	151
Lou Ning, Zheng Xiaobo, Yang Yongchong	Expression model of multi-resolution 3D geographical space	157
Chen Qing, Yong Zhong, Liuming Xiang	Cloud database dynamic route scheduling based on polymorphic ant colony optimization algorithm	161
A Aitmagambetov, Yu Butuzov, S Torekhan	Matters of satellite queuing network design in Ka-band for Republic of Kazakhstan	166

Operation research and decision making		
R Uskenbayeva, B Kurmangaliyeva, N Mukazhanov	Development of multidimensional model of data for information and analytical decision making support system	170
Shangyao Shi, Shiping Zhao, Jun Li	On tracking ability analysis of linear extended state observer for uncertain system	175
Wei Shang, Maozhu Jin, Jie Wang	Using genetic algorithms on service facilities layouts design	180
Wu Shimei, Pei Yulong, Cheng Guozhu	Method of traffic zone division based on spectral graph theory	186
Yanghe Feng, Baoxin Xiu, Zhong Liu	A dynamic optimization model on decision-makers and decision-layers structure (DODDS) in C2-organization	192
Zhao Suxia, Yang Xuemei, Sun Peihong	Research on the method of determining the service scope of civil airport: <i>the case of Chengdu Shuangliu airport</i>	199
Liu Zhusheng	Tourist flow-control study based on utility	205
Liu Caihong, Xiong Wei	The agent-based warning modelling of internal quality risks in supply chain for manufacturer	212
Cong Guodong	A study on the lock-in risk in IT outsourcing projects : <i>the mechanism and the control system</i>	221
Bingjie Li	A game theory analysis: <i>dumping by multinational company and antidumping in China</i>	227
Zhihong Ma, Jianping Chen, Shouzhen Zeng	A new method based on induced aggregation operator and distance measures for fuzzy decision making	232
Bi Jianxin, Lei Lianghai	Analysis on the effectiveness of China's macroeconomic policy based on the modified Mundell-Fleming model during the post-financial crisis period	240
Sheng Yao, Longshu Li	Approximation reductions in an incomplete variable precision multigranulation rough set	250
Wang Hongbao	Cost-sensitive back-propagation neural network for financial distress prediction	259
Shaoren Wang, Zujun Ma, Bochao Zhuang	Fuzzy location-routing problem for emergency logistics systems	265
Peizhe Li, Lirong Jian, Kun Zhang, Shanshan Pei	Evaluation of regional economic development level based on grey clustering and rough set	274
Zhan'an Zhang, Xingguo Cai	Power purchase plan using minimal cost flow	281
Innovative Education		
Tatyana Koycheva	Novel approaches to form the corporate culture in pedagogical universities	286

Authors' Index

290

Cumulative Index

291



Editors' Remarks

Eldorado*by Edgar Allan Poe*

Gaily bedight,
A gallant knight,
In sunshine and in shadow,
Had journeyed long,
Singing a song,

But he grew old-
This knight so bold-
And o'er his heart a shadow
Fell as he found
No spot of ground
That looked like Eldorado.

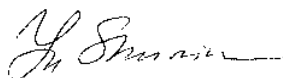
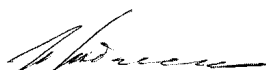
And, as his strength
Failed him at length,
He met a pilgrim shadow-
'Shadow,' said he,
'Where can it be-
This land of Eldorado?'

Over the Mountains'
Of the Moon,
Down the Valley of the Shadow,
Ride, boldly ride,"
The shade replied-
'If you seek for Eldorado!'

Edgar Allan Poe (1809-1849) *

This 18th volume No.2 presents actual papers on main topics of Journal specialization, namely, **Nanoscience and Nanotechnologies. Fundamental Nature Phenomena, Mathematical and Computer Modelling, Information and Computer Technologies, Operation Research and Decision Making and Innovative Education.**

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

* **Edgar Allan Poe** (19 January 1809 - 7 October 1849 / Boston) was an American author, poet, editor and literary critic, considered part of the American Romantic Movement. Best known for his tales of mystery and the macabre, Poe was one of the earliest American practitioners of the short story and is considered the inventor of the detective fiction genre. He is further credited with contributing to the emerging genre of science fiction.



CNT-Fe-Pt interconnect electromagnetic simulations for magnetically stimulated CNT growth and novel memory nanodevices

**Yu N Shunin^{1, 2, 3*}, S Bellucci¹, Yu F Zhukovskii², V I Gopeyenko³,
N Burlutskaya³, T Lobanova-Shunina⁴, A. Capobianchi⁵, F Micciulla¹**

¹ INFN-Laboratori Nazionali di Frascati, Via Enrico Fermi 40, I-00044, Frascati-Rome, Italy,

² Institute of Solid State Physics, University of Latvia, Kengaraga Str. 8, LV-1063 Riga, Latvia

³ Information Systems Management Institute, 1 Lomonosov, Bld. 6, LV-1019 Riga, Latvia

⁴ Riga Technical University, 1 Lomonosov, Bld. 4, LV-1019 Riga, Latvia

⁵ Istituto di Struttura della Materia-CNR A.d.R. Roma, Via Salaria Km 29.300 Monterotondo (Rome) Italy

Received 1 March 2014, www.tsi.lv

Abstract

The parametrically controlled production of carbon nanotubes (CNTs) with predefined morphologies is a topical technological problem for modern nanoelectronics. The chemical vapor deposition (CVD) technique for single walled carbon nanotubes (SWCNTs) in the presence of various metal nanoparticle catalysts is generally used now. The application of a magnetically stimulated CVD process scheme and catalyst nanoparticles with a strong magnetism promises additional possibilities for the CVD process management and allows expecting a predictable growth of CNTs with set chiralities and diameters. The main attention is focused on the magnetically anisotropy Pt-Fe in L10 crystallographic phase nanoparticles effect research. The developed cluster approach based on the multiple scattering theory formalism, realistic analytical and coherent potentials, as well as effective medium approximation (EMA-CPA), can be effectively used for modeling of nanosized systems. It allows us to calculate the dispersion law $E(k)$, electronic density of states, conductivity, etc. This theoretical approach is used for simulation of fundamental electromagnetic properties in Pt-Fe L10-CNT interconnects, which are responsible for developing CNTs morphologies. The developed model of 'effective bonds' and the model of magnetic stimulation for growing CNTs morphologies generated on the Pt-Fe nanoparticle surface are applied for the evaluation of the expected CNT chiralities distribution. The model and conditions controlled magnetically, which stimulate CNT growth in the CVD process, aimed at the predictable SWCNT diameter and chirality and based on Pt-Fe L10 catalyst are discussed. The possibilities of CNT forest growing on FePt nanoparticles for magnetic nanomemory are also evaluated.

Keywords: chemical vapor deposition (CVD), CNTs CVD magnetically controlled growth, arc discharge technique CNTs creation, Pt-Fe nanodrops -catalysts, fundamental electromagnetic properties in Pt-Fe L10-CNT interconnect, model of magnetically controlled CNTs growth

1 Introduction

Carbon nanotubes (CNTs) of various chiralities open new wide possibilities for modern nanoelectronics as promising candidates for nanointerconnects in a high-speed electronic nanosensing and nanomemory devices [1-3].

We focus our current study on the implementation of advanced simulation models for a proper description of the fundamental electromagnetic properties (electrical resistance, capacitances and impedances) in contacts between carbon nanotubes of different morphologies and metallic substrates of different nature. We also present the model of magnetically stimulated CNT growth for a special case of Fe-Pt metallic nanoparticles, which have unique magnetic properties. We expect that in the

presence of magnetic field, the CNTs growth will be more determined from the point of view of possible CNTs morphologies. Moreover, the creation of a CNT forest based on PtFe nanoparticles provides the possibilities to consider this kind of structure as the basic fragment of nanomemory devices, where information bits are located in nanoparticles and the CNT forest provides the necessary spin transport for reading and recording information.

Thus, in our simulations, we expect to reproduce not only a CNT forest with the predefined morphology but also to develop a prototype of a nanomemory device.

The adequate description of CNT chirality [2] is one of the key points for a proper simulation on electric properties of CNT-based nanoelectronic devices.

* Corresponding author- Email: shunin@isma.lv

The electronic structure of CNT-Me interconnects can be evaluated through the electronic density of states (DOS) for a carbon-metal contact considered as a 'disordered alloy', where clusters containing both C and Me atoms behave as scattering centres. The computational procedure developed for these calculations [2, 4] is based on the construction of cluster potentials and the evaluation of both scattering (S) and transfer (T) matrices.

The cluster formalism was successfully applied, e.g., for metallic Cu [4], as well as for both elemental (Ge, Si, As, Se, Te) and binary (As_xSe_{1-x} and Sb_xSe_{1-x}) semiconductors [4, 5]. When using the coherent potential approach (CPA) as EMA approximation, the resistance of interconnect can be evaluated through the Kubo-Greenwood formalism [6] and Ziman model [7].

The further development of the cluster approach allowed us to formulate the 'effective bonds' model and to carry out a cycle of simulations on electromagnetic properties in various CNTs- and graphene (GNRs - graphene nanoribbons) -based metal interconnects (Me = Fe, Cu, Ag, Pt, Au, Pd, Ni) [3, 8]. We consider that this model serves as a tool for understanding the process of CNT growth adequate to CVD process.

2 Research motivation

The main goal of the current research is to understand if there is a relationship between the use of magnetic catalysts and the CVD growth of CNTs, taking into account that most commonly used materials for the growth of CNTs are just Fe, Co and Ni nanoparticles. The nanoparticles of these catalysts are magnetically isotropic. The key question arises: What would happen if one used instead of Fe, Co, and Ni nanoparticles, magnetically anisotropic nanoparticles such as those in the alloy Fe-Pt? The anisotropy of Fe-Pt alloy is due to a spin-orbit coupling of Fe and Pt orbitals. This coupling

takes place only if the alloy is formed of alternating planes of Fe and Pt. This particular structure is called L10. In this way, because of the different sizes of the atoms of Fe and Pt, the structure is crushed and atoms of Fe and Pt are close forming a centred tetragonal phase (fct). This rapprochement allows the coupling and then the magnetic anisotropy with the magnetization axis perpendicular to such planes. It is essential in this case to find conditions to control the growing CNT chirality. We also should take into account a possible substitutional disorder of Fe_xPt_{1-x} alloy, when the stoichiometry number 'x' becomes an additional parameter of CNT growth. No doubt, the diameter of the grown CNT is also an essential parameter, which is evidently pre-defined by the created nanoparticle diameter. But again, we should discuss some limitations in the creation of Fe-Pt nanoparticle sizes in connection with the Fe-Pt melting point. However, Fe-Pt nanoparticles demonstrate an extremely strong coercing field (about, for the phase fct in some special compositions and this promises definite hopes for the controlled CNT growth in the CVD process. We also consider the CVD process of CNT growth as a more predictable one from the point of view of the expected growing CNT parameters (diameter, chirality, morphology of SW or MW CNTs)).

Our experience with the nanotubes synthesis by means of the arc discharge brought about large doubts in the possibilities of CNTs growth control. Two cathodes of pure graphite (99.7%) were taken $\varphi = 6\text{mm}$ (cathode) and $\varphi = 10\text{mm}$ (anode) in diameter. Carbon Black was added to iron - platinum powder, after which the mixture was introduced to complete the filling of the hole. At this point the whole mass was placed in the room. The synthesis was carried out according to the parameters presented here: Current 90 A, Voltage $\sim 18 \div 20\text{V DC}$, Vacuum $\sim 2 \times 10^{-4}$ mbar, Gas: He, Work Pressure 600 mbar, arc duration ~ 15 min.

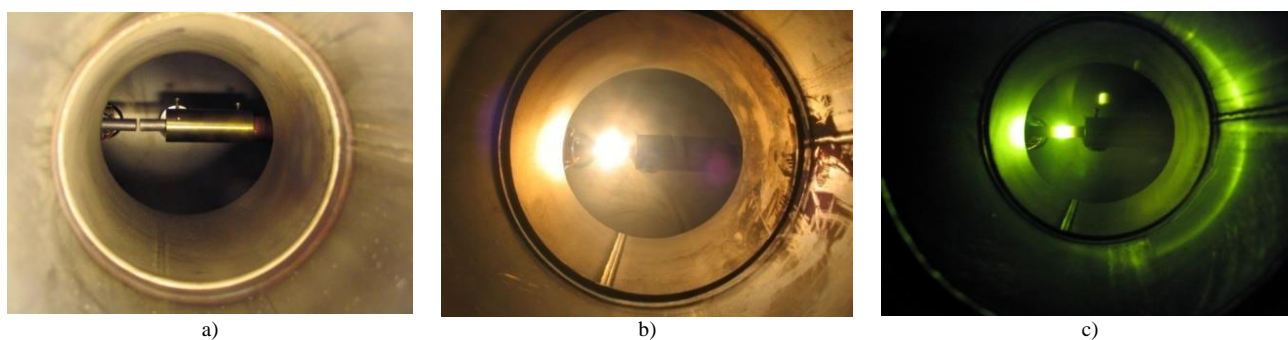


FIGURE 1

a) Cathode and anode aligned before the synthesis

b) Synthesis, arc,

c) arc with a glass front. Once you open the synthesis chamber it is possible to identify the growth of "filamentous" retiring from the anode

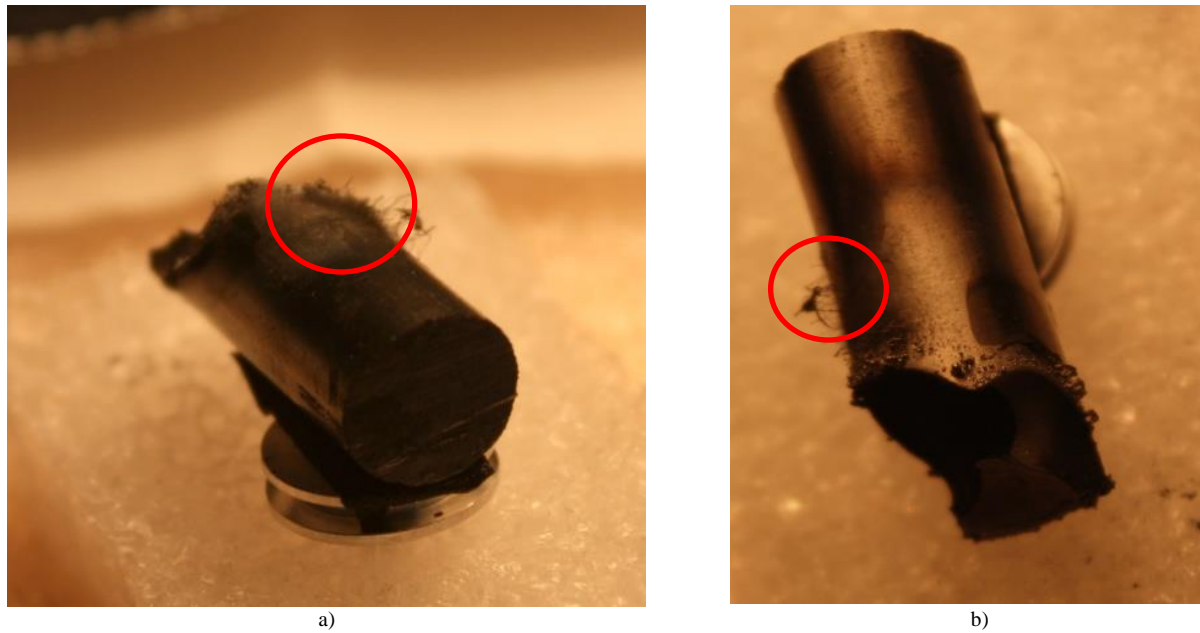


FIGURE 2 Two views of the anode with filaments after exiting out of the synthesis: a); b)

Some filaments also exit from the outer wall of the cathode but in smaller quantities than found on the anode. The residue of the deposit formed on the upper part of the synthesis chamber was collected. It appeared to be quite little compared to other discharges with other catalyst.

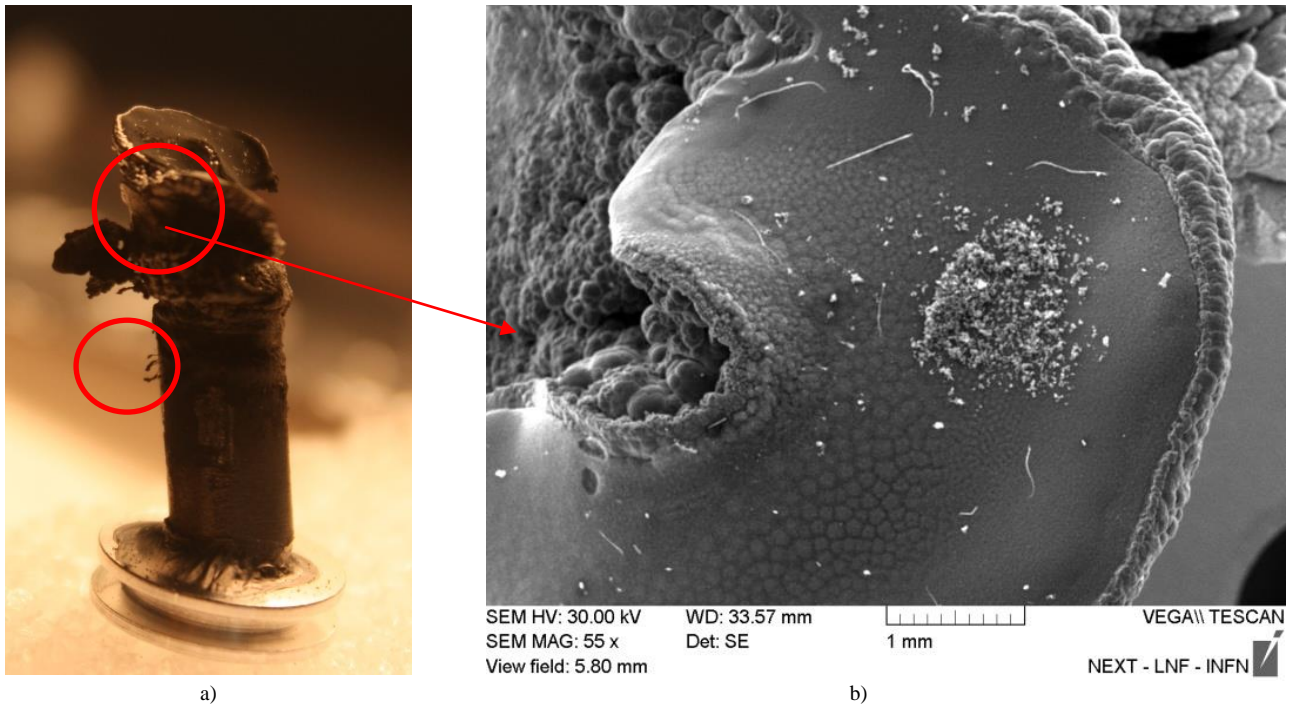


FIGURE 3 a) Photo of the cathode filaments exiting out after the synthesis, b) SEM photo of the cathode surface

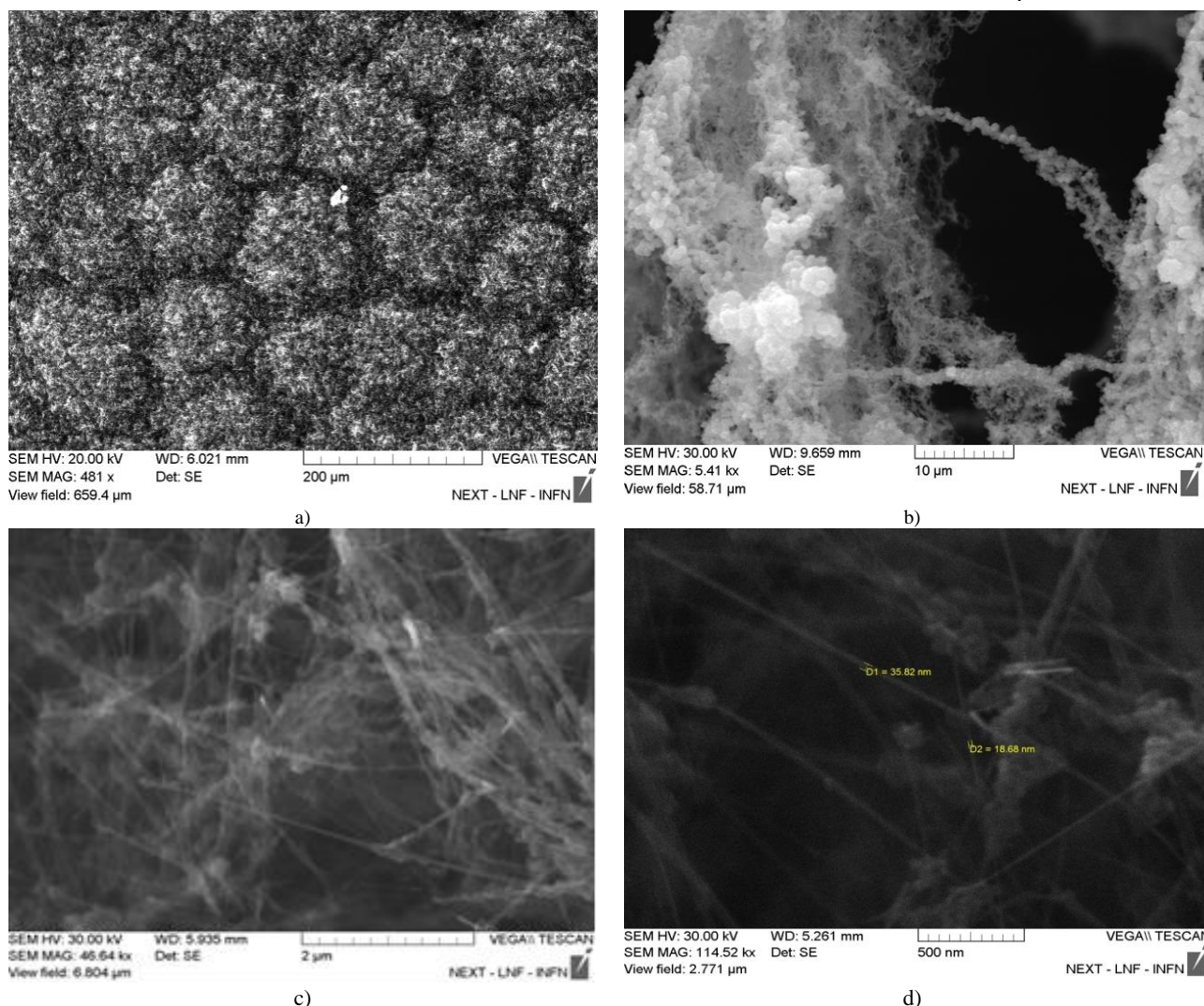


FIGURE 4 Evolution view of filaments in the arc discharge process (Bellucci-Capobianchi): a),b),c),d) [60-63].

The figure 4 shows that filaments are of different nature - some formats are created from these spheres and others have very fine structures of a few nanometers, probably single-walled nanotubes, covered by these spheres, partly or totally. The question is: What percentage of nanoparticles was originally made or how much iron and platinum and how this underwent a change, during the arc discharge process?

We also take into account the extremely marginal parameters of the Pt-Fe system: Pt has the melting point at $T = 2041.4 \text{ K}$ (1768.2°C) and the boiling point at $T = 4098 \text{ K}$ (3825°C);

There are four major crystal structure modifications for Fe: below 769°C (Curie temperature) it is $\alpha\text{-Fe}$ (α -ferrite) with a body-centered cubic crystal structure and ferromagnetic properties. Ferrite above the critical temperature ($769\text{-}917^\circ \text{C}$) is beta-ferrite ($\beta\text{-Fe}$) where it is paramagnetic rather than ferromagnetic and it is crystallographically identical to $\alpha\text{-Fe}$. Within the interval of $917\text{-}1394^\circ \text{C}$ it is $\gamma\text{-Fe}$ (austenite) with acicular cubic crystal structure. At temperatures between 1394°C and

1538°C , the body-centered cubic crystal structure is the more stable form of delta-ferrite ($\delta\text{-Fe}$), the melting point 1538°C (1811 K), and the boiling point at $T = 3273 \text{ K}$ (3000°C).

Our earlier research proved that close to the arc the temperature ranges between $4000 - 6000 \text{ K}$ and there is a subsequent rapid temperature gradient decrease with the distance from it. Filaments are found nearby the arc itself, more precisely, upstream of it. Taking into account the essentially non-steady character of the **arc discharge** process and very high working temperatures, the CNT growth control is practically impossible and CNT morphologies are non-predictable.

3 CNTs growth in the chemical vapor deposition process based on metal nanoparticles

3.1 CVD PROCESS ANALYSIS

The Chemical Vapour Deposition (CVD) is a highly versatile approach to producing nanotubes and is,

perhaps, the most commonly published technique for nanotube growth often used to synthesize CNTs for commercial applications. The process involves decomposition of a carbonaceous precursor at high temperatures under oxygen-free conditions and reduced atmosphere to produce nanotubes.

The carbonaceous precursor may be a carbon gas, methane or ethylene for instance, or it can be in the form of a volatile hydrocarbon solvent such as ethanol, which is generally fed into the reactor with some inert carrier gas. In the reaction chamber, the precursor is decomposed in the presence of a catalyst –usually a transition metal such as iron, cobalt or some similar combination of metals. The catalyst can be introduced in a number of different ways.

The substrate needs to have catalyst nanoparticles providing the growth of CNTs where nanoparticles present the place from which nanotubes start growing. Nanoparticles can be composed of lots of different substances (usually, a metal like Fe, Co, Ni Co, Mo, Mn, Pd etc) [9].

The substrate is then heated to the temperature of about 700-800°C under H₂. As the nanoparticles heat up, the hydrogen pulls oxygen out from the nanoparticles, leaving behind metallic nanoparticles. At the target temperature, a carbon-containing gas such as ethylene (or some others, e.g., alcohol vapor) is introduced. At these temperatures, the gas partly decomposes, giving carbon-containing fragments and other molecules. These

fragments and molecules then work their way onto the catalyst nanoparticles, where they stick and then break down into carbon. The carbon dissolves with some probability into or on the nanoparticle. When a critical concentration of carbon is reached in the nanoparticle, the addition of just a little bit more carbon from the vapor causes carbon to precipitate as a crystal.

When carbon crystallizes under the atmospheric pressure (like in CVD), it crystallizes into flat sheets of hexagonally-patterned carbon. However, because the nanoparticle is so small and has such a high curvature, when the carbon crystallizes out, it is constrained into a cylindrical shape in the form of a carbon nanotube. Additional carbon atoms on the nanoparticle provide further nanotube growth, which stops when the carbon-containing gas is turned off if the substrates cool down, or for a number of other reasons that can influence the nanotube or catalyst nanoparticles.

Tuning the appropriated parameters, it is possible to obtain a huge amount of CNTs in a cheap and quick way. In addition, it is the technique enabling the growth of CNTs with specific characteristics: particular diameter, length and orientation (alignment), and family (Multi or Single Wall). It has been well known for a long time that carbon nanotubes are synthesized by catalytic decomposition of hydrocarbon [10] in the reactor, where metal nanoparticles are presented as a catalyst on a substrate. Catalyst metals mostly used for these purposes are listed in the following Table 1 [11, 12].

TABLE 1 Most commonly used catalyst metals for the CVD method [11].

Catalyst			Temperature, °C	Carbon source
Metal	Catalyst type	Preparation method		
Fe	Ultra fine particle	Decomposition of metallocene	1060	Benzene
	Silica support	Pore impregnation	700	Acetylene
	Zeolite or Clay support	Ion exchange	700	Acetylene
	Graphite support	Impregnation	700	Acetylene
	Ultra fine particle	Decomposition metal of carbonyl	800	Acetylene
	Silica support	Sol-gel process	700	Acetylene
Co	Ultra fine particle	Laser etching of Co thin film	1000	Triazine
	Ultra fine particle	Decomposition metal of carbonyl	800	Acetylene
	Silica support	Pore impregnation	700	Acetylene
	Zeolite or Clay support	Ion exchange	700	Acetylene
	Graphite support	Impregnation	700	Acetylene
Ni	Graphite support	Impregnation	700	Acetylene
	Ultra fine particle	Decomposition of Ni(C ₈ H ₁₂) ₂	800	Acetylene
Mo	Ultra fine particle	Decomposition of Mo* ¹	800	Acetylene
Mn	Ultra fine particle	Decomposition metal of carbonyl	800	Acetylene
Pd	Ultra fine particle	Decomposition metal of carbonyl	800	Acetylene

¹Mo*=(NH₄)₂₅₋₅₅[Mo₁₅₄(NO)₁₄O₂₂₀(OH)₂₈(H₂O)₇₀] 350H₂O.

Hydrocarbon, such as methane, adsorbed on the catalytic particle surface, releases carbon during decomposition, which dissolves and diffuses into a metal particle. When a supersaturated state is reached, carbon precipitates in a crystalline tubular form.

It is necessary to emphasize that it will be possible to grow CNTs using a gas, both as a catalyst and as hydrocarbon. If a stream of catalyst particles can be

injected into the flowing feedstock, it is possible to produce nanotubes in the gas phase. This approach is amenable for scale-up to large-scale production. Sen et al. [13] first reported such a possibility when they used ferrocene or nickolecene as a source of the transition metal and benzene as a carbon source. This approach yielded MWNTs, whereas their later work [14] with gas-phase pyrolysis of acetylene using a metallocene yielded

SWNTs with diameters around 1 nm [15]. CVD-produced CNTs are curved and have high amount of defects, mainly after the purification in the acid bath; it is needed to remove metal nanoparticles inside of tubes.

The main interests related to iron-containing nanoparticles are focused on their potential applications as high-quality magnetic materials. Thus, a new generation of iron-metal nanoparticles is studied to be used as a catalyst in growing CNTs, with a magnetic nanomaterial inside, making it possible to design a precise carbon nanomagnetic device for drug delivery diagnostics.

For example, iron nanoparticles have been widely used as a catalyst for CVD synthesis of multiwalled carbon nanotubes, [16] while iron-molybdenum can act as a very efficient catalyst for the synthesis of either single-walled or multiwalled carbon nanotubes with the CVD method [17-22].

To prevent the magnetic behaviour of these iron – nano-compounds, it is necessary to grow carbon nanotubes of nanocompounds below the Curie temperature. Typically, the growth temperature is around 600 – 700°C at the atmospheric pressure [23].

The potential applications of carbon nanotubes grown for semiconductor and sensor devices are presented for

CMOS industrial applications [24]. The low-temperature growth of vertically aligned carbon nanotubes (CNTs) at high growth rates by a photo-thermal chemical vapour deposition (PTCVD) technique using a Ti/Fe bilayer film as a catalyst is presented in [25]. The bulk growth temperature of the substrate is as low as 370 °C and the growth rate is up to 1.3 μmmin^{-1} , at least eight times faster than the values reported by traditional thermal CVD methods.

It should be recognized that the mechanism of CNTs growth is not obvious enough. Technological growth characteristics may differ in detail, although the concept remains the same. The mechanism of carbon atoms deposition on metal catalyst nanoparticles with the subsequent nucleation of CNTs can be considered one of the most effective and practically important. Crystal (eg. Si) nanoparticles with a set diameter created on a substrate give a high probability of producing CNTs with the regulated diameter. However, the problem of controlling the chirality of CNTs remains a pressing one. Among more effective catalysts, it is possible to distinguish Pt, Pd, Cu, Ag, Au, Si, SiC, Ge, Al₂O₃ [26] in addition to the earlier investigated Mg, Ti, Cr, Mn, Fe, Co, Ni, Sn, Pb [27-29].

TABLE 2 Comparison of catalysts for SWCNTs growth [27-29]

Type	State	Size for SWCNT growth	Catalysis for cracking	Carbon solubility	Growth temperature for ethanol (°C)	Growth rate
Fe, Co, Ni	Liquid/ solid	<10 nm	Yes	High	600-950	High
Pt, Pd	Liquid/ solid	<5 nm	No?	High	850-950	High
Au, Ag, Cu	Liquid/ solid	<5 nm	No	Low	850-950	High
Ge	Liquid/ solid	<5 nm	No	Low?	850-950	Low?
Si, SiC	Solid	<5 nm	No	No?	850-950	Low
Al ₂ O ₃	Solid	<5 nm	No	No	850-950	Low

3.2 ADVANTAGES OF CVD

Compared to the discharge and laser-ablation methods, CVD is a simple and economic technique for synthesizing CNTs at low temperatures and ambient pressures. Arc- and laser-grown CNTs are superior to the CVD-grown ones. In yield and purity, CVD has advantages over the arc and laser methods. Considering CNTs in relation to structure or architecture control, CVD is the only answer. CVD is versatile in the sense that it offers harnessing plenty of hydrocarbons in any state (solid, liquid or gas), enables the use of various substrates, and allows CNT growth in a variety of forms, such as powder, thin or thick films, aligned or entangled, straight or coiled nanotubes, or a desired architecture of nanotubes on predefined sites of a patterned substrate. It also offers better control over the growth parameters zone, while carbon crystallization (being an endothermic process) absorbs some heat from the metal precipitation zone. This precise thermal gradient inside the metal particle keeps the process going on. This last advantage is

the most essential one [26]. Taking only two main CNT parameters, namely, CNT diameter and chirality, the problem cannot be solved well enough. There are two marginal cases of CVD growth.

- **Tip-growth model** – the catalyst–substrate interaction is weak (metal has an acute contact angle with the substrate); hydrocarbon decomposes on the top surface of the metal; carbon diffuses down through the metal, and CNT precipitates out across the metal bottom, pushing the whole metal particle off the substrate; as long as the top of the metal is open for fresh hydrocarbon decomposition, CNT continues to grow longer and longer; the metal is fully covered with the excess carbon and its catalytic activity ceases and the CNT growth stops.
- **Base-growth model** – the catalyst–substrate interaction is strong; initial hydrocarbon decomposition and carbon diffusion take place similar to that in the tip-growth case; the CNT precipitation fails to push the metal particle up, so the precipitation is compelled to emerge out of the metal apex; carbon

crystallizes out as a hemispherical dome, which then extends up in the form of seamless graphitic cylinder; subsequent hydrocarbon deposition takes place on the lower peripheral surface of the metal, and a dissolved carbon diffuses upward; CNT grows up on the catalyst nanoparticle base.

3.3 CNT PRECURSORS

Most commonly used CNT precursors are methane [30-31], ethylene [14, 17] acetylene [32], benzene [13] xylene [33] and carbon monoxide [34].

SWCNTs were first produced from the disproportionation of carbon monoxide at 1200 °C, in the presence of molybdenum nanoparticles [35], and later they were produced from benzene [36] acetylene [14], ethylene [37], methane [38], cyclohexane [39], fullerene [40] by using various catalysts.

It means that the working temperature and pressure of the CVD process can be changed in a wide range of ways and the kinetics of carbon atoms deposition can also be widely varied.

In 2002, the low-temperature synthesis of high-purity SWCNTs from alcohol on Fe–Co-impregnated zeolite support was carried out [41] and since then, ethanol has become the most popular CNT precursor in the CVD method worldwide [42-44].

A special interest in binary catalysts becomes principal now. The unique feature of ethanol is explained by the fact that ethanol-grown CNTs are almost free from amorphous carbon, owing to the etching effect of OH radical [45]. Later, vertically-aligned SWCNTs were also grown on Mo-Co-coated quartz and silicon substrates [46-47]. It has been shown that intermittent supply of acetylene in ethanol CVD significantly assists ethanol in preserving the catalyst activity, thus, enhancing the CNT growth rate [48].

Generally, low-temperature CVD (600-900° C) yields MWCNTs, whereas high-temperature (900–1200 °C) reaction favours SWCNTs growth. It means that SWCNTs have a higher energy of formation (presumably owing to small diameters and the fact that high curvature bears high strain energy). Probably, this is the reason that MWCNTs are easier to grow than SWCNTs.

Recent developments in the nanomaterials synthesis and characterization have enabled many new catalysts for the CNTs growth. Apart from popularly used transition metals (Fe, Co, Ni), a range of other metals (Cu, Pt, Pd, Mn, Mo, Cr, Sn, Au, Mg, Al) have also been successfully used for horizontally-aligned SWCNT growth on quartz substrates [49]. It has been also proposed that the active catalyst is Au–Si alloy with about 80 at% Au [50].

3.4 CNT GROWTH CONTROL

It is a general experience that the catalyst-particle size dictates the tube diameter. The particle size dependence

and a model for iron-catalysed growth of CNTs has been reported in [51]. Metal nanoparticles of the controlled size, pre-synthesized by other reliable techniques, can be used to grow CNTs of the controlled diameter [52].

3.4.1 Influence of a catalyst material and concentration

The additional advantage of using the bimetallic catalyst is that CNTs can be grown at a much lower temperature - 550°C. E.g., the melting point of the mixture of Fe and Co is lower than their individual melting points. Moreover, alloys are known to be better catalysts than pure metals. These trends suggest that tri-metallic catalysts should also give interesting results, though the interpretation of the results would be more complicated.

3.4.2 Influence of temperature

There are investigations of the temperature effect on camphor CVD in a wide range of temperatures 500-1000°C [53]. It was noticed that camphor did not decompose below 500 °C. At 550 °C very short-length tubes emerged from the zeolite pores suggesting that the catalyst activity, and hence the CNT growth rate, was quite low at 550°C. However, the CNT growth abruptly increased at 600 °C, and a profound growth was observed all around the zeolite pores. At 650°C and above, the growth rate was so enormous that hardly a zeolite particle could be located amid nanotubes. The CNT diameter is increased with the growth temperature increase. Very pure CNTs, almost free from metallic impurity, were produced up to 750°C. From 750°C onward, both the diameter and the diameter-distribution range increased drastically. It is supposed that at high temperature, the metal atoms agglomerate into bigger clusters leading to thick CNTs. At 850°C and above, SWCNTs began to take shape alongside with MWCNTs and the volume of SWCNTs increased with the increasing temperature. For instance, at 900 °C, samples of large bundles of SWNTs can be observed. The CVD temperature plays the central role in CNT growth. For a fixed metal concentration, the increasing CVD temperature enlarges the diameter distribution. It should also be noted that MWCNTs and SWCNTs can be selectively grown as a function of CVD temperature if the catalyst concentration is properly optimized [53, 54].

3.4.3 Influence of pressure

For the controlled growth of CNTs by CVD, the vapor pressure of the hydrocarbon in the reaction zone is another very important parameter. For gaseous hydrocarbons, a desired vapor pressure in the CVD reactor can be maintained by a limited gas-flow rate and the controlled suction with a rotary pump [55]. In the case of a liquid hydrocarbon, its vapor pressure is controlled by its heating temperature before it enters the reactor

[56]. However, for a solid hydrocarbon such as camphor, it is quite problematic to control its vapor pressure. It becomes a function of three parameters: camphor mass, its vaporization temperature, and the flow rate of argon—the carrier gas. By proper optimization of these three parameters, the influx of camphor vapor to the zeolite bed and its decomposition rate were balanced to a great extent, and a record growth of MWCNTs was achieved at atmospheric pressure by CVD [9]. Pure SWCNTs (free from MWCNTs) have been selectively obtained from camphor CVD at low pressures (10–40 torr) where the camphor vapor pressure is quite in tune with the low metal concentration [53]. It is important to note that the melting point of nanoparticles below 10 nm falls abruptly. E.g., a 5-nm Fe and Co particles can melt at about 850 °C and 640 °C, respectively [9]. It leads to essential limitations in the CVD process of small diameter (< 5nm) CNTs growth.

Schematic representation of the basic steps of SWCNT growth on a metal catalyst is usually simulated in three steps [58]:

- 1) the diffusion of single C atoms on the surface of a catalyst;
- 2) the formation of an sp^2 graphene sheet floating on the catalyst surface with edge atoms covalently bonded to the metal;
- 3) the root incorporation of diffusing single C atoms. It has been shown that carbon atoms diffuse only on the outer surface of the metal cluster. At first, a graphene cap is formed which floats over the metal, while the border atoms of the cap remain anchored to the metal. Subsequently, more C atoms join the border atoms pushing the cap up and, thus, constituting a cylindrical wall.

4 Magnetically stimulated CNTs CVD growth on Fe-Pt catalysts

4.1 EFFECTIVE BONDS MODEL FOR CNT-Pt-Fe INTERCONNECTS ELECTROMAGNETIC PROPERTIES

We have developed structural models for CNT-Me and GNR-Me junctions, based on their precise atomistic structures - clusters, which take into account the CNT chirality effect and its influence on the interconnect resistance for Me (= Fe, Ni, Cu, Ag, Pd, Pt, Au) as well as the pre-defined CNT (or GNR) geometry. These atomistic structures are in compliance with the proposed

‘effective bonds’ model. The ‘effective bonds’ are responsible for mechanical, electronic, magnetic and electrical properties of interconnects. The common consideration of two marginal carbon structures (CNTs and GNRs) is induced by the similar technological problems in respect of these materials for the modern nanoelectronics.

Our main interest in this particular case is focused entirely on CNT-Me interconnects, when the modelling of CNTs growth on metal surfaces is combined with the controlled electromagnetic properties in the interconnect area. To ensure a predictable interconnect morphology, the permanent magnetic field is used that accompanies the growth, and magnetic drops are used as growth catalysts that provide for the special composition of Fe-Pt with unique magnetic properties.

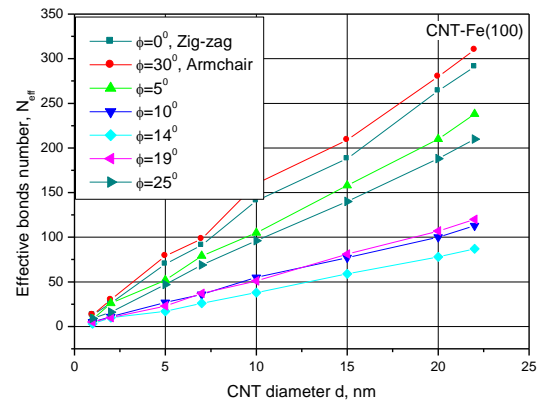
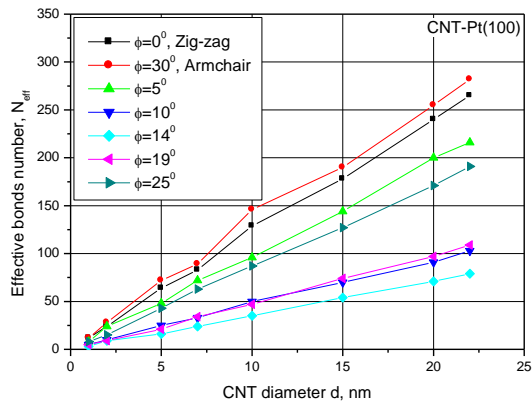
The results of our simulations show that interconnects resistance and the number of effective bonds can be considered as indicators of chirality.

Figures 5, 6, 7 and 8 demonstrate the numbers of effective bonds via CNT diameters, chirality angles, CNT-Me interconnects resistances and impedances..

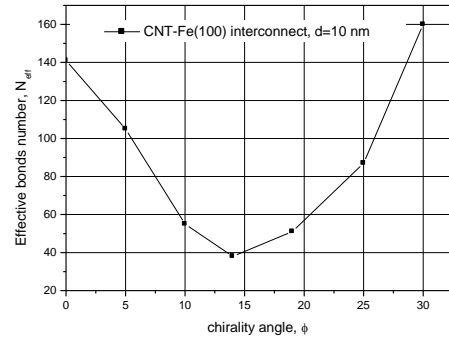
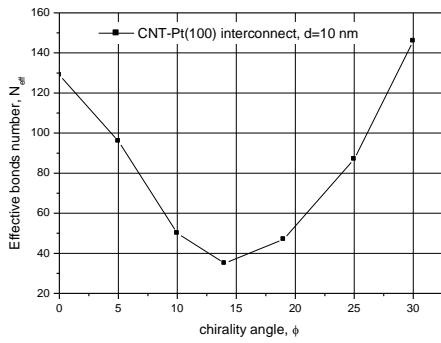
It means that resistances and the number of effective bonds in the interconnect space are indicators of CNTs morphology.

4.2 CNT- Fe_xPt_{1-x} INTERCONNECT FORMATION

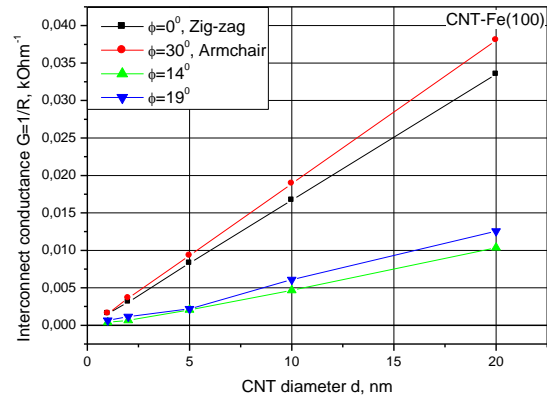
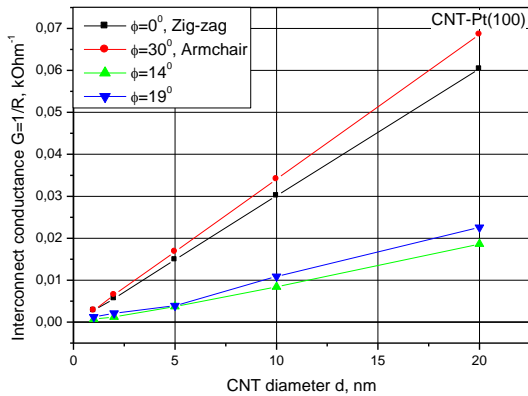
There is a relation between the use of magnetic catalysts and the CVD growth of CNTs determining the most commonly used materials for the CNT growth - Fe, Co and Ni. The nanoparticles of the latter catalysts are magnetically isotropic. Magnetism of the particles-catalysts brings order into the process of CVD growth. Additional possibilities in controlling the process of CVD growth open up the ways to use magnetically anisotropic nanoparticles such as those in the alloys with a different substitutional disorder (e.g., Fe_xPt_{1-x}), to manage the CVD process with the formation of the predefined CNT chiralities. Besides, the stimulation of the process by means of the external magnetic field activates magnetic moments of the catalyst and the deposited carbon atoms. This also means that we can control the number of effective bonds inside interconnects. We pay a special attention to Pt and Fe substrates as possible elements for nanoelectronic and nanomagnetic devices considering interconnects fundamental properties and magnetically stimulated nanoprocesses on Fe-Pt $L1_0$ nanoparticles [59].



a) b)
FIGURE 5 Effective bonds number via CNT diameter: a) CNT-Pt; b) CNT-Fe.



a) b)
FIGURE 6 Effective bonds number via CNT chirality for the diameter $d=10$ nm: a) CNT-Pt; b) CNT-Fe.



a) b)
FIGURE 7 Interconnect conductance via CNT diameter: a) CNT-Pt; b) CNT-Fe.

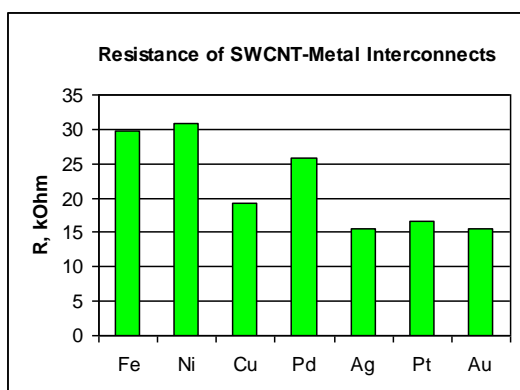


FIGURE 8 Resistances of zig-zag SWCNT-Metal interconnects in comparison.

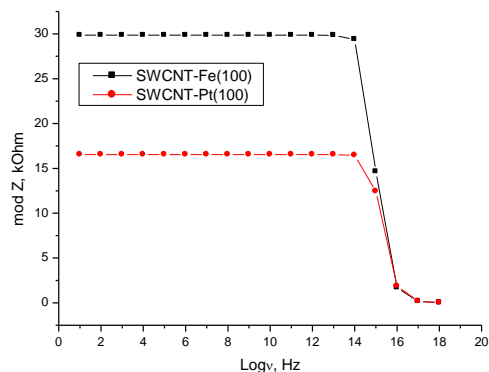


FIGURE 9 Impedances of SWCNT-Fe and -Pt interconnects in comparison.

It means that resistances and the number of effective bonds in the interconnect space are indicators of CNTs morphology.

4.2 CNT-Fe_xPt_{1-x} INTERCONNECT FORMATION

There is a relation between the use of magnetic catalysts and the CVD growth of CNTs determining the most commonly used materials for the CNT growth - Fe, Co and Ni. The nanoparticles of the latter catalysts are magnetically isotropic. Magnetism of the particles-catalysts brings order into the process of CVD growth. Additional possibilities in controlling the process of CVD growth open up the ways to use magnetically anisotropic nanoparticles such as those in the alloys with a different substitutional disorder (*e.g.*, Fe_xPt_{1-x}), to manage the CVD process with the formation of the predefined CNT chiralities. Besides, the stimulation of the process by means of the external magnetic field activates magnetic moments of the catalyst and the deposited carbon atoms. This also means that we can control the number of effective bonds inside interconnects. We pay a special attention to Pt and Fe substrates as possible elements for nanoelectronic and nanomagnetic devices considering interconnects fundamental properties and magnetically stimulated nanoprocesses on Fe-Pt L1₀ nanoparticles [59].

4.2.1 Magnetic properties of Fe-Pt alloys.

The unique magnetic properties of Fe-Pt alloys - is an open field for research to correlate different CNT techniques in terms of the catalyst role in entirely different range of temperatures and pressures. Carbon

nanotubes grow using bimetallic nanoparticles Fe-Pt as a catalyst [64-67]. Since the mid-1930s Fe-Pt alloys have been known to exhibit high coercivities due to high magnetocrystalline anisotropy of the L1₀ FePt phase but their high cost prevented these alloys from widespread applications in the past. In Fe-Pt alloys, both Fe and Pt atoms carry a magnetic moment: The induced magnetic moment on the Pt sites and the enhanced magnetic moment on the Fe sites. A wide variety of the magnetic structure types in the Fe-Pt alloys is evidently the consequence of various atomic configurations around Fe atoms, which, in turn, has a considerable effect on the electronic structure of these alloys (see, Figure 10, and Table 3)

TABLE 3 Specific magnetic structures of Pt-Fe systems

Magnetic Structure	Behaviour	Curie Temperature
L1 ₂ Fe ₃ Pt	Ferromagnetic	410K
L1 ₀ FePt	Ferromagnetic	750K
L1 ₂ FePt ₃	Paramagnetic	273K

To obtain the Fe-Pt alloy in L1₀ phase we use a new chemical synthesis strategy reported in details in our previous works [60-63]. By planetary ball milling of the nanocrystals of a particular precursor [Fe(H₂O)₆]PtCl₆ and NaCl were grounded and then annealing at 400 °C (300 °C less than typical used temperature) in reductive atmosphere. After washing we obtain FePt L1₀ NPs with selected size as function of fraction of [Fe(H₂O)₆]PtCl₆ and NaCl used. By varying the precursor ratio we are able to avoid the coalescence phenomena and obtain single crystal NPs with the size around 6 nm, not agglomerated.

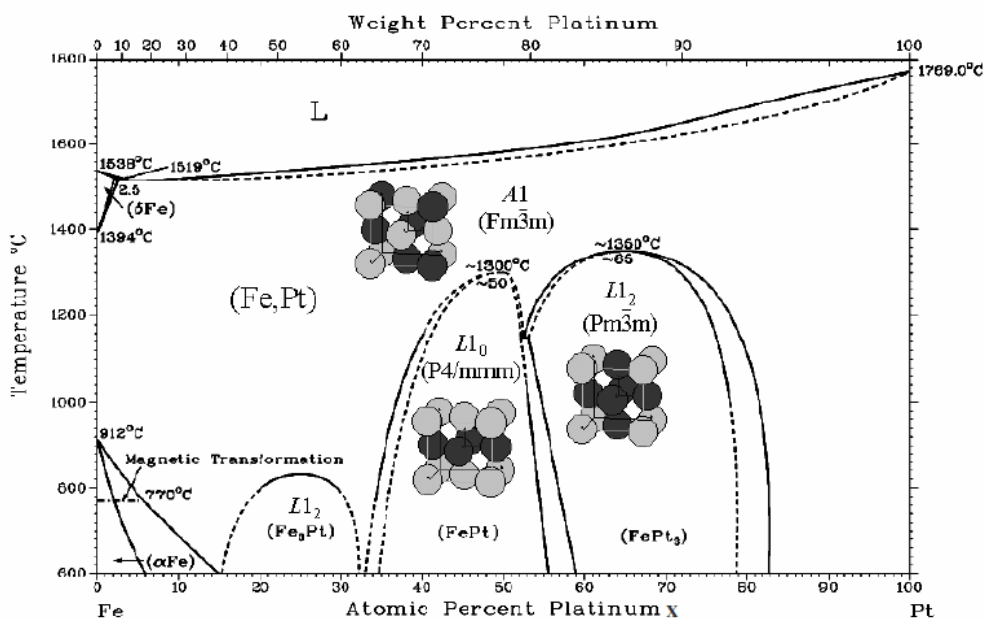


FIGURE 10 Eutetic Phase Diagram of Fe-Pt alloy (<http://www.himikatus.ru/art/phase-diagr1/Fe-Pt.php>)

4.2.2 Magnetically stimulated CNTs growth with the probabilistically pre-defined morphology

The formation of the initial optimal perimeter for C-Metal (Fe-Pt) bonds is a synergetic process with a minimal free energy (see Figure 11).

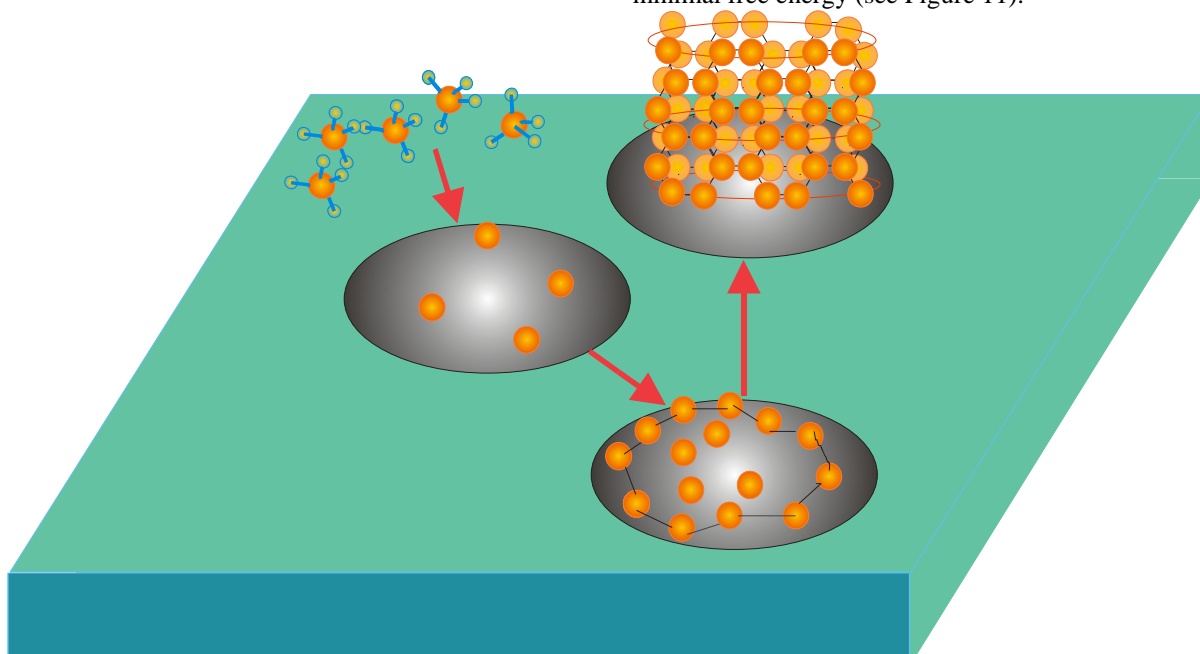


FIGURE 11 A fragment of the CNT CVD process growth on nanoparticle surfaces

The nanoparticle diameter determines with a certain error the diameter of a CNT. The number of effective bonds defines the morphology of the future CNT (See Fig.12, arm-chair, chiral, zig-zag CNTs,) in terms of chirality.

Obviously, there is a considerable uncertainty in the morphology of the future CNTs, owing to sporadic thermal dynamics of the deposited carbon atoms.

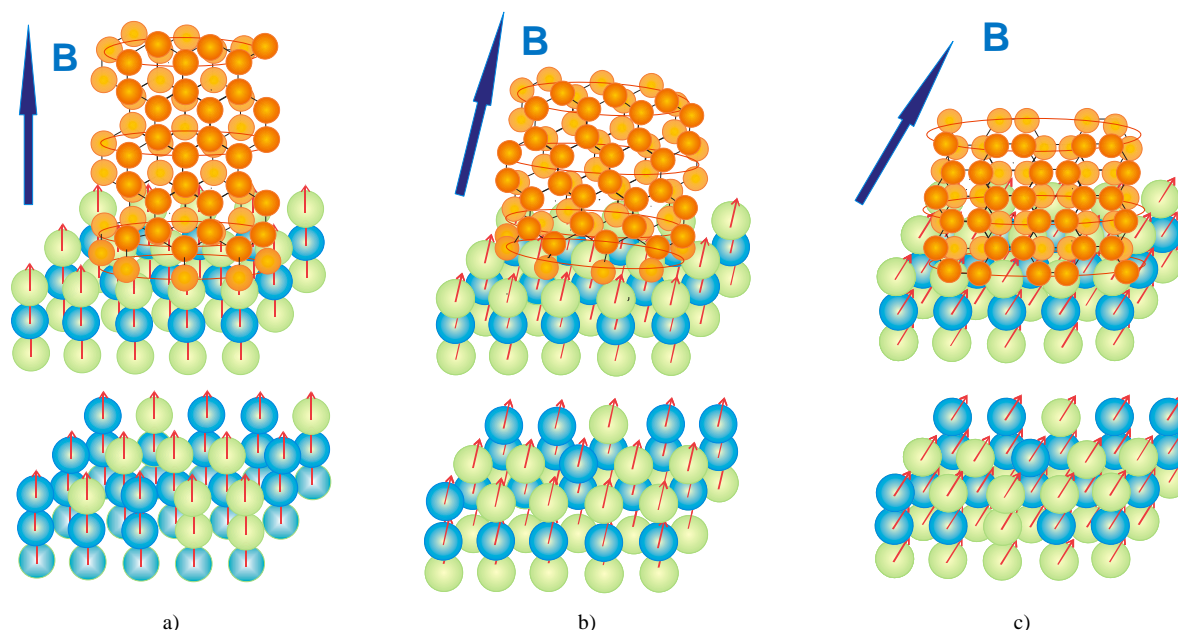


FIGURE 12 Magnetically stimulated orientation of magnetic moments of atoms in model interconnects during the CNTs controlled growth with the expected chiralities under the directed mag $\phi = \Theta_B$: a) arm-chair CNT, netic field \mathbf{B} with the orientation angle Θ_B , where \mathbf{B} is the magnetic induction and the chirality angle is $\phi = 0^\circ$; b) chiral CNT, $0^\circ < \phi < 30^\circ$; c) zig-zag CNT, $\phi = 30^\circ$

According to our model, the process can be considerably streamlined using even minor diamagnetic properties of carbon atoms at the expense of magnetic field and strong induced ferromagnetism of the nanoparticle. It becomes evident that the nanoparticle composition Fe-Pt and its atomic structure (including the short-range order) play a primary role in the process of CNT growth stimulated by the magnetic field of CVD.

Nucleation in the process of CNT growth in cases of the ordered Fe-Pt nanoparticles is more stable and has principal advantages in relation to the controlled CNTs growth compared with the cases of any kind of anisotropic Fe-Pt nanocatalysts [68].

The availability of substitutional disorder and anisotropy in Fe-Pt nanodrops used for CNT CVD controlled growth is a negative factor for the production of CNTs with predictable morphologies. There are some essential negative features of disorder in Fe-Pt nanodrops for the CVD process:

- 1) C atoms chemical binding preferences (C-Fe bonds are stronger than C-Pt ones),
- 2) A higher structural anisotropy leads to higher relaxation times of a nanodrop structure [69]
- 3) Disorder effects in Fe-Pt lead to sporadic magnetization phenomena

Thus, it is important to control the process of formation of nanodrop structures corresponding to $L1_0$ (Fe-Pt) and $L1_2$ (Fe_3Pt) systems with reliable magnetism and adequate cooling rates, when relaxation processes in nanodrops are accomplished.

4.2.3. Model of CVD CNT growth

Magnetically stimulated CNT CVD model formulation looks as follows (see Figs.11, 12):

- Effective bonds are responsible for the future chirality of CNT in the CVD processes.
- The permanent magnetic field provides the alignment of CVD growth.
- The binding with the Fe-Pt nanoparticle has a probabilistic character; the probability of binding of C atoms with Fe atoms is preferable as compared to Pt atoms.
- The composition of Fe-Pt atoms in a nanoparticle and the atomic structure ordering are essential for chirality of the obtained CNTs.
- The diameter of the future CNT is correlated with the nanoparticle size.
- The direction of the magnetic field stabilizes the direction of CNT forest growth and is a tool of the growth control.
- The ordering of a nanoparticle atomic structure, as well as a substitutional disorder or arbitrary disorder of a nanoparticle, is responsible for the magnetic state of the nanoparticle, which is essential for the possible future nanomemory devices.

Even a very strong magnetic field is not able to suppress thermal fluctuations of magnetic moments in carbon atoms. It is a different thing, when carbon diamagnetic atoms appear on the surface of a catalyst in the ferromagnetic condition where magnetism induced by the external field can be very strong and is able to correct

the behaviour of carbon atoms by suppressing thermal fluctuations.

At the same time, the growth control over chiral and non-chiral nanotubes essentially depends on stoichiometric composition of Pt-Fe nanoparticles. The beginning of the nucleation process with the growth of nanotubes might be connected with stochastic fluctuations of the magnetic moment in a carbon atom relative to the direction of the local magnetic field in a nanoparticle. Distribution of the fluctuation angle obeys the Gaussian law: $f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\theta-\theta_B)^2}{2\sigma^2}\right)$, where

σ^2 is the angular dispersion of thermal fluctuations of the magnetic moment angle of a carbon atom. To evaluate this dispersion, the potential energy change of the magnetic moment under the influence of the thermal energy should be evaluated: $\mu_C B(1 - \cos \theta_T) \approx k_B T_{CVD}$, where μ_C is the induced magnetic moment of a carbon atom $\mu_C = 1.25\mu_B$ (see evaluations in [65], $\mu_B = 5.788 \cdot 10^{-5} eV/T$, \mathbf{B} is the magnetic induction of the catalyst surface, $\theta_T = \theta - \theta_B$, T_{CVD} - is the operating temperature of the CVD process, $k_B = 8,617 3324(78) \cdot 10^{-5} eV/K$ - is the Boltzmann constant. Hence $2 \sin^2 \frac{\theta_T}{2} \approx \frac{k_B T_{CVD}}{\mu_C B}$. Taking into consideration one of the main problems of the nanotubes growth control – the chirality control – it is necessary to seek for the small fluctuation angle θ . Then $\sigma^2 = \theta_T^2 = \frac{2k_B T_{CVD}}{\mu_C B}$.

The condition of the small fluctuation angle (e.g., $<10^\circ$) at a certain temperature of the CVD process imposes limitations on the values of the demanded magnetic induction \mathbf{B} .

Taking into account the ratio between the chirality angle and the direction of the magnetic field $\phi = \theta_B$, Figure 13 displays the predictable scattering of chiralities for nanotubes of approximately the same diameter.

We are also able to evaluate the necessary value of the magnetic field \mathbf{B} providing the expected chirality angles scattering, e.g. $\sigma = 0.2$ (approximately 12°) leads to the \mathbf{B} evaluation for the CVD process temperature $T_{CVD} = 700^\circ C$ as

$$B = \frac{2k_B T_{CVD}}{\sigma^2 \mu_C} \approx \frac{16755}{\sigma^2 \mu_C} \approx 57895 \approx 6 \cdot 10^4 T. \text{ For small}$$

angle dispersions $\sigma^2 = \theta_T^2 = \frac{2k_B T_{CVD}}{\mu_C B}$ the high local

magnetic field on the nanoparticle surface is necessary. The result also strongly depends on the carbon atom magnetic moment μ_C .

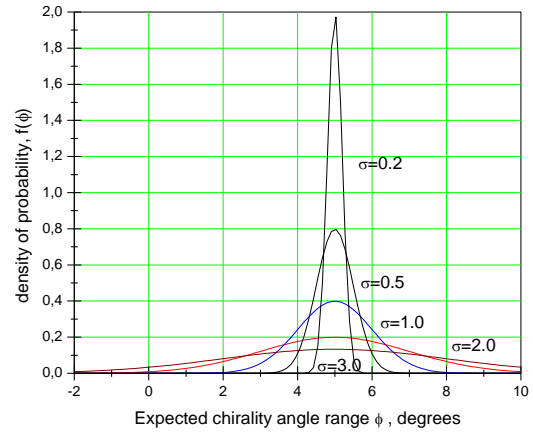


FIGURE 13 The predictable scattering of chiralities for nanotubes of approximately similar diameters

Speaking about the possible errors in diameters of growing nanotubes, their evaluation from beneath is defined by the minimal variations in parameters of the chirality vector $\vec{c} = (n, m)$ Δn and Δm , which are equal 1.

Taking into consideration the formula for calculating the diameter of CNT: $d = \frac{\sqrt{3}a}{\pi} \sqrt{m^2 + n^2 + mn}$, where $a = 0,142$ nm is the distance between the neighbouring carbon atoms in the graphite plane.

The relationship between chirality indices (n and m) and the angle ϕ is given by the ratio:

$$\sin \phi = \frac{m\sqrt{3}}{2\sqrt{m^2 + n^2 + mn}}.$$

Then the minimum relative error in the diameter of CNT is possible to define as:

$$\varepsilon_d = \frac{\Delta d}{d} \approx \frac{\sqrt{5m^2 + 5n^2 + 8mn}}{2(n^2 + m^2 + mn)}.$$

Particularly, in the case of arm-chair CNT ($m=0$) $\varepsilon_d \approx \frac{\sqrt{5}}{2n}$, in the case of zig-zag CNT ($m=n$)

$$\varepsilon_d \approx \frac{1}{2\sqrt{2}n}.$$

Thus, the errors in the diameter of growing CNTs are incorporated in discrete morphological properties. However, these minimum estimates are only reinforced, given the obvious errors in the size of catalyst nanoparticles are taken into consideration.

The perfect picture of the magnetically stimulated CVD process for growing CNTs can be presented as a CNT forest (see Figure 14). Such a system of nanotubes can also be considered as a prototype of the magnetic

memory, where ferromagnetic nanoparticles serve as cells of the magnetic memory – that is, ferromagnetic contacts

are controlled by spin pulses, the transport of which is provided by nanotubes.

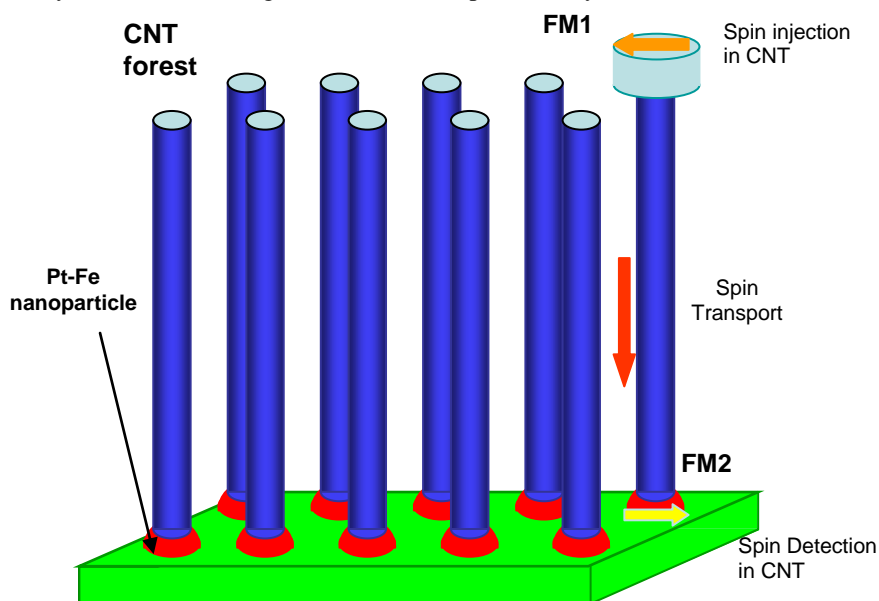


FIGURE 14 CNTs forest is grown on the Fe-Pt nanoparticles of the predefined radius as a fragment of magnetic nanomemory device, which can be realized if spin injection, spin transport and spin detection (spin recording) are provided

5 Conclusions

Electromagnetic properties of Pt- and Fe-CNT interconnects are considered from the point of view of mechanical stability and electrical efficiency. CNT-Fe interconnects are stronger mechanically. However, CNT-Pt interconnects having smaller resistances are more suitable electrically and more effective for various electronic nanodevices.

The use of Pt-Fe nanoclusters during the CVD magnetically stimulated growth allows controlling with some limitation the diameter and chirality of growing CNTs. The diameter of nanoclusters defines the diameter of CNTs, while the angles of chiralities are correlated with the orientation of external permanent magnetic field.

Pt-Fe nanoclusters and CNTs composition is the prototype of the magnetic nanomemory devices. The magnetic efficiency depends on the stoichiometry coefficient 'x' and the ordering of Pt-Fe atomic structure.

Chemically ordered Fe-Pt nanoparticles, where we meet the sequence of Pt and Fe layers, allows us to provide all possible predicted magnetic properties. In case of substitutional disorder, this advantage disappears and we should talk about the percolation phenomenon of Fe_xPt_{1-x} ferromagnetism via 'x' alteration.

The CVD process of CNTs growth with the presence of Fe-Pt nanoparticles in the conditions of a strong magnetic field is a more orderly process.

Statistical dispersion of CNTs output according to the diameter is determined by a discrete number of effective bonds (especially in the case of small diameters of growing CNTs).

Here with, the dependence of the number of the effective bonds, formed on the nanoparticle perimeter, on the chirality angle is the essential tool for the morphology control of the future CNT.

The balance factor of the thermal energy and magnetic energy in deposited nanoparticles of carbon atoms plays the decisive role in evaluating the dispersion of chirality angles in CNTs.

The morphology of the forthcoming growing CNT stimulated by the magnetic field is set when forming the perimeter of the CNT base and the character of the magnetic orientation in effective bonds.

We can consider CNT forest magnetically stimulated by the CVD process as a fragment of the magnetic nanomemory. CNT chirality in this case defines the efficiency of magnetic cells access (rates of exchange).




Acknowledgments

This study was supported by Grant EU FP7 CACOMEL project FP7-247007, Call ID 'FP7-PEOPLE-2009-IRSES', 2010-2014 *Nanocarbon based components and materials for high frequency electronics*. Evaluating the contribution of the research team members we should point out that V I Gopeyenko and N Yu Burlutskaya made a large amount of calculations, Yu F Zhukovskii and F Micciulla made the critical analysis of CVD details of CNTs growth, Yu N Shunin, T Lobanova-Shunina, and S Bellucci made the review of theoretical approaches and wrote most of the manuscript. We also thank Prof. E A Kotomin and Prof. S A Maksimenko for stimulating discussions on the topic.

References

- [1] Shunin Yu, Kiv A, ed 2012 *Nanodevices and Nanomaterials for Ecological Security. Series: Nato Science for Peace Series B - Physics and Biophysics* Hiedelberg: Springer Verlag 363p
- [2] Shunin Yu N, Zhukovskii Yu F, Burlutskaya N Yu, Gopejenko V I, Bellucci S 2012 In: *Nanodevices and Nanomaterials for Ecological Security, Series: Nato Science for Peace Series B - Physics and Biophysics*, ed Yu Shunin and A Kiv Hiedelberg: Springer Verlag 237-262
- [3] Shunin Yu N, Zhukovskii Yu F, Gopejenko V I, Burlutskaya N, Lobanova-Shunina T and Bellucci S 2012 *Journal of Nanophotonics* **6**(1), 061706-1-16
- [4] Shunin Yu N, Schwartz K K 1997 *Computer Modelling of Electronic and Atomic Processes in Solids* ed R C Tennyson and A E Kiv Dodrecht/Boston/London: Kluwer Acad. Publisher 241-257
- [5] Shunin Yu N 1991 *Dr.Sc.Habil. Thesis (Phys&Math)* Rigas-Salaspils: Institute of Physics
- [6] Economou E L (2006) *Green's Functions in Quantum Physics Solid State Ser. 7* (3rd edition) Berlin/Heidelberg: Springer Verlag
- [7] Ziman J M 1979 *Models of Disorder* New York-London: Cambridge Univ. Press Ch 10
- [8] Shunin Yu N, Zhukovskii Yu F, Burlutskaya N, Bellucci S 2012 *Journal of Nanoelectronics and Optoelectronics* **7**(1) 3-11
- [9] Moiala A, Nasibulin A G, Kauppinen E I 2003 *J. Phys. Condens. Mater.* **15**, S3011-35
- [10] Baker R T K, Hams P S 1978 *Chemistry and Physics of Carbon* **14** ed P L Walker Jr and P A Thrower New York: Marcel Dekker 83p
- [11] Muller T E, Reid D G, Hsu W K, Hare J P, Kroto H W, Walton D R M 1997 *Carbon* **35** 951
- [12] *The Science and Technology of Carbon Nanotubes* 1999 ed K Tanaka, T Yamabe, K Fukui Amsterdam: Elsevier Science Ltd
- [13] Sen R, Govindaraj A, Rao C N R 1997 *Chem.Phys.Lett.* **267** 276
- [14] Satishkumar B C, Govindraj A, Sen R, Rao C N R 1998 *Chem.Phys.Lett.* **293** 47
- [15] *Carbon Nanotubes Science and Applications* 2005 ed M Meyyappan Florida: CRC Press LLC Google e-Book
- [16] Xie S, Chang B, Li W, Pan Z, Sun L, Mao J, Chen X, Qian L, Zhou W 1999 *Adv.Mater.* **11** 1135
- [17] Cassell A M, Raymakers J A, Kong J, Dai H J 1999 *Phys.Chem. B* **103** 6484
- [18] Su M, Zheng B, Liu J 2000 *Chem.Phys.Lett.* **322** 32
- [19] Su M, Li Y, Maynor B, Buldam A, Lu J P, Liu J J, *Phys.Chem. B* **104** 6505
- [20] Fan S S, Chapline M G, Franklin N R, Tomblor T W, Cassell A M, Dai H 1999 *Science* **283** 512
- [21] Cheng H M, Li F, Su G, Pan H Y, He L L, Sun X, Dresselhaus M S 1998 *Appl.Phys.Lett.* **72** 3282
- [22] Li Y, Liu J 2001 *Chem. Mater.* **13**, 1008-14
- [23] Lee D C, Mikulec F V, Korgel B A 2004 *Amer. Chem. Soc.* **126** 4951
- [24] Chen G Y, Jensen B, Stolojan V, Silva S R P 2011 *Carbon* **49**(1) 280-5
- [25] Shang N G, Tan Y Y, Stolojan, V, Papakonstantinou P, Silva S R P 2010 *Nanotechnology* **21**, 505604
- [26] Kumar M, Ando Y 2010 *Journal of Nanoscience and Nanotechnology* **10** 3739
- [27] Takagi D, Hibino H, Suzuki S, Kobayashi Y, Homma Y 2007 *Nano Lett.* **7** 2272
- [28] Takagi D, Homma Y, Hibino H, Suzuki S, Kobayashi Y 2006 *Nano Lett.* **6** 2642
- [29] Liu H, Takagi D, Ohno H, Chiashi Sh, Chokan T, Yo. Homma Yo 2008 *Appl.Phys.Express* **1** 014001
- [30] Hernadi K, Fonseca A, Nagy J B, Bernaerts D, Lucas A A, 1996 *Carbon* **34** 1249
- [31] Kong J, Cassell A M, Dai H 1998 *Chem. Phys. Lett.* **292** 567
- [32] Li W Z, Xie S, Qian L X, Chang B H, Zou B S, Zhou W Y, Zhao R A, Wang G 1996 *Science* **274** 1701
- [33] Wei B Q, Vajtai R, Jung Y, Ward J, Zhang R, Ramanath G, Ajayan P M 2002 *Nature* **416** 495
- [34] Nikolaev P, Bronikowski M J, Bradley R K, Rohmund F, Colbert D T, Smith K A, Smalley R E 1999 *Chem. Phys. Lett.* **313** 91
- [35] Dai H, Rinzler A G, Nikolaev P, Thess A, Colbert D T, Smalley R E 1996 *Chem. Phys. Lett.* **260** 471
- [36] Cheng H M, Li F, Sun X., Brown S D M., Pimenta M A, Marucci A, Dresselhaus G, Dresselhaus M S 1998 *Chem. Phys. Lett.* **289** 602
- [37] Hafner J H, Bronikowski M J, Azamian B R, Nikolaev P, Rinzler A G, Colbert D T, Smith K A, Smalley R E 1998 *Chem. Phys. Lett.* **296** 195
- [38] Flahaut E, Govindaraj A, Peigney A, Laurent C, Rao C N R 1999 *Chem. Phys. Lett.* **300** 236
- [39] Gruneis A, Rummeli M H, Kramberger C, Grimm D, Gemming T, Barreiro A, Ayala P, Pichler T, Kuzmany H, Schamann C, Pfeiffer R, Schumann J, Buchner B 2006 *Physica Status Solidi B* **243** 3054
- [40] Maruyama S, Miyauchi Y, Edamura T, Igarashi Y, Chiashi S and Murakami Y 2003 *Chem. Phys. Lett.* **375** 553
- [41] Maruyama S, Kojima R, Miyauchi Y, Chiashi S, and Kohno M, 2002 *Chem.Phys.Lett.* **360** 229
- [42] Okubo S, Sekine T, Suzuki S, Achiba Y, Tsukagoshi K, Aoyagi Y and Kataura H, 2004 *Jpn. J. Appl. Phys.* **43** L396
- [43] Gruneis A, Rummeli M H, Kramberger C, Barreiro A, Pichler T, Pfeiffer R, Kuzmany H, Gemming T, Buchner B 2006 *Carbon* **44** 3177
- [44] Nasibulin A G, Moiala A, Jiang H, Kauppinen E I 2006 *J. Nanopart. Res.* **8** 465
- [45] Murakami Y, Miyauchi Y, Chiashi S, Maruyama S 2003 *Chem. Phys. Lett.* **377** 49
- [46] Murakami Y, Chiashi S, Miyauchi Y, Hu M, Ogura M, Okubo T and Maruyama S 2004 *Chem. Phys. Lett.* **385** 298
- [47] Maruyama S, Einarsson E, Murakami Y, Edamura T 2005 *Chem. Phys. Lett.* **403** 320
- [48] Xiang R, Einarsson E, Okawa J, Miyauchi Y and Maruyama S. 2009 *J.Phys.Chem.C* **113** 7511
- [49] Yuan D, Ding L, Chu H, Feng Y, McNicholas T P, Liu J 2008 *Nano Lett.* **8** 2576
- [50] Ward J, Wei B Q, Ajayan P M 2003 *Chem.Phys.Lett.* **376** 717
- [51] Morjan R E, Nerushev O A, Sveningsson M, Rohmund F, Falk L K L, Campbell E E B 2004 *Appl. Phys. A* **78** 253
- [52] Ago H, Komatsu T, Ohshima S, Kuriki Y, Yumura M 2000 *Appl.Phys.Lett.* **77** 79
- [53] Kumar M, Ando Y 2005 *Carbon* **43** 533
- [54] Li W Z, Wen J G, Ren Z F 2002 *Appl. Phys. A* **74** 397
- [55] Li W Z, Wen J G, Tu Y, Ren Z F 2001 *Appl. Phys. A* **73** 259
- [56] Maruyama S, Murakami Y, Shibata Y, Miyauchi Y, Chiashi S 2004 *J. Nanosci. Nanotechnol.* **4** 360
- [57] Kumar M, Ando Y 2008 *Defence Sci. Journal* **58** 496
- [58] Raty J-Y, Gygi F, Galli G 2005 *Phys.Rev.Lett.* **95** 096103
- [59] Bellucci S, Zhukovskii Yu F, Gopeyenko V I, Burlutskaya N, Shunin Yu N 2012 *CIMTEC2012 Theses*, June 10-14 Montecatini, Terme, Tuscany, Italy, 78
- [60] Capobianchi A, Colapietro M, Fiorani D, Foglia S, Imperatori P, Laureti S, Palange E 2009 *Chem. Mater.* **21**(10) 2007
- [61] Capobianchi A, Campi G, Camalli M, Veroli C 2009 *Z. Kristallogr.* **224** 384
- [62] Faustini M, Capobianchi A, Varvaro G, Grosso D 2012 *Chem. Mater.* **24** 1072
- [63] Bellucci S. 2005 *physica status solidi* (c) **2**(1) 34
- [64] Sun A C, Kuo P C, Chen S C, Chou C Y, Huang H L, Hsu J H 2004 *J. Appl. Phys.* **95**(11) 7264
- [65] Elkins K, Li D, Poudyal N, Nandwana V, Jin Zh, Chen K, Liu J P 2005 *J. Phys. D* **38** 2306
- [66] Yan M L, Sabirianov R F, Xu Y F, Li X Z, Sellmyer D J 2004 *IEEE Trans. on Magnetism* **40**(4) 2470
- [67] Kim Y-H, Choi J, Chang K J 2003 *Phys. Rev. B* **68** 125420
- [68] Medwal R, Sehdev N, Annapoorani S 2012 *J. Phys. D Appl. Phys.* **45** 055001
- [69] Arabshahi H, Hematabadi A, Bakhshayeshi A, Ghazi M 2012 *World Applied Programming* **2** (8) 415-20

Authors	
	<p>Yuri N Shunin</p> <p>Current position, grades: Professor and Vice-Rector on academic issues at Information Systems Management University and a leading researcher at the Institute of Solid State Physics, University of Latvia.</p> <p>University studies: PhD (physics and maths, solid state physics, 1982) at the Physics Institute of Latvian Academy of Sciences and Dr. Sc. Habil. (physics and maths, solid state physics, 1992) at Saint-Petersburg Physical Technical Institute (Russia).</p> <p>Scientific interest: His current research activities concern nanophysics, nanoelectronics, nanodevices, nanomaterials, nanotechnologies, nanorisks, nanoeducation, and nanothinking</p> <p>Publications: over 470, 1 book with Springer</p> <p>Experience: director of NATO ARW "Nanodevices and Nanomaterials for Ecological Security" Riga, Latvia, 2011, a visiting researcher at Gesellschaft für Schwerionenforschung mbH, Darmstadt, Germany (1995), INFN—Laboratori Nazionali di Frascati, Frascati-Roma, Italy (2010 to 2014), participation in EU FP7 Projects CATHERINE (2008 to 2011) and CACOMEL (2010 to 2014), experience in Higher Education from 1975 till nowadays</p>
	<p>Stefano Bellucci</p> <p>Current position, grades: PhD, Professor, currently coordinates all theoretical physics activities at INFN Laboratori Nazionali di Frascati (Italy).</p> <p>University studies: April 1982: Laurea in Physics (Magna cum Laude), University of Rome "La Sapienza" July 1984: Master in Physics of Elementary Particles, SISSA and University of Trieste, PhD in physics of elementary particles in 1986 at SISSA, Trieste, Italy.</p> <p>Publications: over 400 papers in peer-reviewed journals (with h = 40), and more than 10 invited book chapters, the editor of ten books with Springer</p> <p>Scientific interest: research interests include theoretical physics, condensed matter, nanoscience and nanotechnology, nanocarbon-based composites, and biomedical applications.</p> <p>Experience: He worked as a visiting researcher at the Brandeis University, Waltham, MA, USA (1983 to 1985); at the M.I.T., Cambridge, MA, USA (1985 to 1986); the University of Maryland, USA (1986 to 1987); at the University of California at Davis, USA (1987 to 1988). He is an editorial board member of the Springer Lecture Notes in Nanoscale Science and Technology, as well as the editorial board member of the Global Journal of Physics Express and the Journal of Physics & Astronomy</p>
	<p>Yuri Zhukovskii</p> <p>Current position, grades: Dr.Chem, Head of Laboratory of Computer Modelling of Electronic Structure of Solids, Institute of Solid State Physics (University of Latvia).</p> <p>Publications: He is the author of over 120 regular and review papers in international scientific journals. His Hirsch index is 16.</p> <p>Experience: From 1977 until 1995 he was a researcher at the Institute of Inorganic Chemistry, Latvian Academy of Sciences. Since 1995 he has been a leading researcher at the Institute of Solid State Physics, University of Latvia. Within the last 20 years he has been granted several fellowships for collaboration, visiting activities and positions at seven universities and scientific centres of Canada, Finland, Germany, United Kingdom, and the United States. He has also been actively engaged in developing active collaboration with some scientific groups in Belarus, Italy, Russia, and Sweden. Simultaneously, he has been a contact person and participant in a number of collaboration projects under support of European Commission. His current research activities concern theoretical simulations on the atomic and electronic structure of crystalline solids (with 3D, 2D and 1D dimensionalities).</p>
	<p>Victor Gopeyenko</p> <p>Current position, grades: professor, Dr.Sc.Eng Vice-Rector on scientific issues at Information System Management University and the director of ISMU Computer Technologies Institute.</p> <p>University studies: Riga Civil Aviation Engineering Institute (Latvia), obtained his doctor's degree (Dr. Sc. Eng., 1987) at Riga Civil Aviation Engineering Institute (Latvia).</p> <p>Scientific interest: current research activities concern nanophysics, nanoelectronics, nanodevices, and nanotechnologies in the EU FP7 Project CACOMEL (2010 to 2014). His special interests concern carbon nanotubes and graphene systems applications and modelling.</p> <p>Publications: 80 regular papers in international scientific journals</p> <p>Experience: he was the member of local organizing committee of NATO ARW "Nanodevices and Nanomaterials for Ecological Security," Riga, Latvia, 2011, he is the editor-in-chief of the journal Information Technologies, Management and Society and editorial board member of the journal Innovative Information Technologies.</p>
	<p>Nataly Burlutskaya</p> <p>Current position, grades: a researcher at the Information Systems Management University and the Institute of Solid State Physics, University of Latvia</p> <p>University studies: Master degree in computer systems (2011) at Information Systems Management University, Riga, Latvia.</p> <p>Scientific interest: current research activities concern theoretical simulations of the electronic and electrical properties of carbon nanotubes and graphene nanoribbons in the EU FP7 Project CACOMEL (2010 to 2014).</p> <p>Publications: 30 regular papers</p> <p>Experience: Nataly Burlutskaya was the secretary of organizing committee of NATO ARW "Nanodevices and Nanomaterials for Ecological Security," Riga, Latvia, 2011.</p>

	<p>Tamara Lobanova-Shunina</p> <p>Current position, grades: Associate Professor at Riga Technical University, PhD, Dr.edu University studies: University of Latvia, She obtained her PhD (2009) on innovative education at South-Ukrainian National University Scientific interest: current research activities concern nanotechnologies, nanomanagement, nanoeducation, nanorisks, and nanotechnology in the EU FP7 Project CACOMEL (2010 to 2014), special interests concern the systemic approach to nanosystems applications. Publications: 55 regular papers Experience: She was a member of the NATO ARW Local Organizing Committee 'Nanodevices and Nanomaterials for Ecological Security', Riga, Latvia, 2011. Her She has been working as a visiting researcher at INFN-Laboratori Nazionali di Frascati, Frascati-Roma, Italy (2010 to 2014). She was the Head of International Business Communications Department, Director of the study programme 'International Business Communications' at Information Systems Management University (till 2013). She is the Editorial Board member of the journal 'Innovative Information Technologies'. She is the author of over 53 scientific papers in international scientific journals.</p>
	<p>Aldo Capobianchi</p> <p>Current position, grades: Istituto di Struttura della Materia ISM, Dr chem. University studies: Dr Degree in Chemistry (Rome , 1996) taken at CNR <u>Consiglio Nazionale delle Ricerche</u> since 1989, http://www.ism.cnr.it Scientific interest: Inorganic Chemistry, Mechanical Engineering, Material Science Publications: about 30 Experience: Extensive experience in summary: Sitesi and characterization of organometallic compounds to the charge transfer as phthalocyanines and porphyrins. Synthesis and characterization of nanostructured hybrid materials (nanocomposite) based on carbon nanotubes and nanoparticles. (a patent) Synthesis and characterization of nanostructured magnetic materials for magnetic recording. Application experience in various fields: Growth of thin films of organic and inorganic substances with different devices (LB, thermal evaporation, etc.), for applications in the fields of chemical sensors and organic photovoltaics. Chemical gas sensors for environmental pollutants (a patent). Construction and testing of organic solar cells based on phthalocyanines.</p>
	<p>Federico Micciulla</p> <p>Current position, grades: Contrator Research at Frascati National Laboratories - National Institute of Nuclear Physics University studies: Università degli Studi di Roma 'La Sapienza' Aerospace Engineer, Postdoctoral Research Fellow presso Università degli studi di Cassino e del Lazio Meridionale Roma, Italia Scientific interest: Nanotechnology, Polymers, Carbon Nanotubes, Graphene Publications: about 40 Experience: Synthesis of Carbon Nanotubes (CNTs) by Arc Discharge method. Growth of Graphene Nano - Plates (GNPs) by Microwave assistance. Development and project of nano composites as shielding coating against electromagnetic interference, using several Carbon Materials: CNTs, GNPs, Graphene, Carbon Black, Exfoliated Graphite, Activated Graphite. Growth of Carbon Nanotubes inside pores of alumina membrane for novel high speed interconnections for electronics Characterization of nano materials by scanning electron and atom force microscope. Responsible for analysis service and maintenance of s Scanning Electron Microscopes and Atomic Force Microscopes. Interpretation of data, problem resolution</p>

On the classical roots of the Schroedinger equation

A Ershkovich*

Department of Geophysics and Planetary Sciences, Tel Aviv University, Tel Aviv, 69978 Israel

Received 1 March 2014, www.tsi.lv

Abstract

In semiclassical approximation Schroedinger equation is known to reduce to classical Hamilton-Jacobi equation. These equations look strikingly similar. An idea that just the Hamilton-Jacobi equation became a prototype of the Schroedinger equation arises. Arguments in favour of this assumption are supplied. Then it is no wonder that Aharonov-Bohm effect was recently derived directly from the classical Hamilton-Jacobi equation (without using Schroedinger equation), and hence, it is, in fact, of classical origin. The electron-field interaction is explained within the framework of classical electrodynamics. Thus, the so-called unlocal interaction becomes unnecessary.

Keywords: Schroedinger equation, Hamilton-Jacobi equation, Aharonov-Bohm effect

1 Introduction

Speaking on Quantum Mechanics we will mean Erwin Shroedinger and Werner Heisenberg mechanics created in 1925-1926 rather than Quantum Physics or Theory founded by Max Planck in 1901. An analogy between classical and quantum mechanics is known to exist. Some physical effects, for instance, the normal Zeeman effect allow both quantum and classical treatment (by means of magnetic field effect on harmonic oscillations of the atomic electrons). One may believe that each physical analogy should have a mathematical description. For instance, optics-mechanical analogy follows from similarity between Fermat and Maupertuis (or Hamilton) principles. An analogy between Hydrodynamics and Electrodynamics is described by similarity between Helmholtz equation for the vorticity, $\mathbf{\Omega} = \nabla \times \mathbf{v}$:

$$\frac{\partial \mathbf{\Omega}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{\Omega}) \text{ and Faraday equation in the form}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) \text{ for the magnetic field } \mathbf{B}.$$

In contrast to the examples above, the existing analogy between the *Quantum* and *Classical* Mechanics has not been yet expressed mathematically. The basic Schroedinger equation (equivalent to the Heisenberg description) is postulated (or explained) rather than being derived [1]. It is known that Erwin Schroedinger was looking for adequate mathematics which should be able to describe all the relevant experiments. He found it: this is the Schroedinger equation for the wave function Ψ , which governs the Quantum Mechanics:

$$-i\hbar \frac{\partial \Psi}{\partial t} + \hat{H}\Psi = 0, \quad (1)$$

where \hat{H} is the Hamilton operator. However, what is its analogue in classical mechanics?

It is hard to believe that the equation he was looking for dawned upon him instantly. We will never know it for sure but let us assume that the prototype of the equation (1) was the classical Hamilton-Jacobi equation for the action S governing the classical mechanics and electrodynamics

$$\frac{\partial S}{\partial t} + H = 0. \quad (2)$$

There are a number of arguments in favour of this assumption. Formal similarity between equations (1) and (2) is too striking to be accidental. Ignoring the imaginary number i , they formally coincide with $S = \hbar\Psi$ (and hence $H = \hbar\omega$ and *vice versa*). Similarity retains if the Hamiltonian H includes the electromagnetic field [2]. Equation (1) satisfies all the requirements. It is of the first order in time, otherwise, the causality principle would not hold: the time evolution of a system is determined by the

Taylor series: $\Psi(t_0 + \Delta t) = \Psi(t_0) + \left. \frac{\partial \Psi}{\partial t} \right|_{t=t_0} \cdot \Delta t$ for any

moment (chosen to be initial), Δt being infinitesimal in order to ignore higher derivatives.

In other words, the equation has to include Ψ and $\frac{\partial \Psi}{\partial t}$.

It has to be linear in order to satisfy the superposition principle, necessary, for instance, to describe motion of

* Corresponding author- E-mail:alexer@tau.ac.il

atomic electrons. Of course, equation (2) is nonlinear as the classical Hamiltonian depends on quadratic term proportional to $(\nabla S)^2$. But it is easy to overcome this obstacle: replacement of the physical quantities by corresponding operators makes the equation linear as $(\nabla S)^2$ becomes $\nabla^2 S$. There is another difficulty with the equation (2): solutions of partial differential equations of the first order in time describe only irreversible processes (like diffusion or heat transfer) whereas it is necessary to consider also periodic processes like motion of atomic electrons. Therefore E. Schrodinger *ad hoc* inserted the imaginary unit i into his equation in order to get complex solutions as well. Because of that, equation (1) cannot be derived from the known principles or equations.

In semiclassical (or WKB) approximation, when

$$\Psi = a(\mathbf{r}, t) \exp(iS/\hbar). \quad (3)$$

Schrodinger equation (1) is known to reduce to the classical Hamilton-Jacobi equation (2) [1, 3] if the second-order term proportional to \hbar^2 is neglected. In one's turn, in the same short wavelengths limit (called also the approximation of geometrical optics) equation (2) reduces to eikonal equation of the wave optics

$$\frac{\partial \psi}{\partial t} + \omega = 0. \quad (4)$$

This equation sometimes is also called the Hamilton-Jacobi equation (see, e.g. [4]). Here ψ is the phase (eikonal), ω is the wave frequency. Comparing equations (2) and (4) and using the relation $H = \hbar\omega$ for the particle energy one obtains $S = \hbar\psi$. Thus, the eikonal $\psi = S/\hbar$ may be considered as the phase of the de Broglie wave Ψ (3), with the frequency $\omega = H/\hbar$ and wave vector $\mathbf{k} = m\mathbf{v}/\hbar$. We arrive at the conclusion that Hamilton-Jacobi equation (2) describes the wave-particle duality in *classical* physics. Notice that this is the only such an equation in classical physics. Of course, the roots of this duality are located in the close affinity between Fermat and Hamilton principles mentioned above.

2 On the Aharonov-Bohm effect

The Aharonov-Bohm effect was predicted in 1959 [5] in the semiclassical approximation (3). However, just in the same approximation of small wavelengths Schrodinger equation (1) reduces to the classical equations (2) and (4). This fact allows us to arrive at the same result as Aharonov and Bohm [5] without using equation (1) [6, 7].

One obtains the relation $S = \hbar\psi$ from comparison between classical equations (2) and (4) and using the relation $H = \hbar\omega$, which, by the way, has been known long before the creation of Quantum Mechanics in 1926.

On the other hand, the action S for the system of the electric charge q in the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ is known to acquire an additional term S_{in} [8]

$$S_{in} = q \int \mathbf{A} d\mathbf{r}, \quad (5)$$

due to interaction between the charge q and the field \mathbf{B} . As $\psi = S/\hbar$, the phase shift $\Delta\psi$ accumulated due to charge-field interaction equals

$$\Delta\psi = \frac{q}{\hbar} \int \mathbf{A} d\mathbf{r}. \quad (6)$$

Thus, we arrived at the Aharonov-Bohm effect without using the Schrodinger equation (1).

No wonder that the Aharonov-Bohm effect has classical roots. As mentioned above, normal Zeeman effect also has these roots. Its explanation by means of H. A. Lorentz electron theory holds already more than a century, along with the explanation via Quantum Mechanics. These facts may be used as additional indication of classical roots of Quantum Mechanics.

It is worth mentioning that Aharonov-Bohm effect is a source of ideas on the special role of the magnetic potential \mathbf{A} in Quantum Mechanics and so called electron-solenoid unlocal interaction. E. Feinberg [9] was the first to show that the interaction between the electron current $\mathbf{j} = q\mathbf{v}$ and the solenoid, due to Faraday induction has to be taken into consideration. As a result of this interaction, the magnetic field arises outside the solenoid, and the Lorentz force $q\mathbf{v} \times \mathbf{B} \neq 0$. We believe that the very fact that the interaction term S_{in} in equation (5) results in the correct phase shift in equation (6) shows that the classical term S_{in} in the classical Hamiltonian already includes the field arising outside the solenoid due to Faraday induction. Indeed, Maxwell equations (which include the Faraday induction law) may be derived from Hamilton principle of least action S where the charge-field interaction is described by the term S_{in} (equation (5)) [8]. Thus, an assumption on the unlocal interaction becomes unnecessary.

As to special role of potentials (\mathbf{A}, φ) , Feinberg [9] underwent very thorough search but did not find any signs of their special role in experiments suggested in [5]. No wonder, because the Aharonov-Bohm effect was derived in the semiclassical approximation (3) when equation (1) reduces to classical equation (2). Then, where such a role in Quantum Mechanics arises from? The Hamilton function for the electric charge q in the electromagnetic field with potentials (\mathbf{A}, φ) in classical electrodynamics is [8]

$$H = \frac{1}{2m} (\nabla S - q\mathbf{A})^2 + q\varphi, \quad (7)$$

whereas in Quantum Mechanics it is [3]

$$\hat{H} = \frac{1}{2m} (\hat{P} - q\mathbf{A})^2 + q\varphi, \quad (8)$$

where the momentum operator $\hat{P} = -i\hbar\nabla$. By using the Heisenberg uncertainty principle along with the Coulomb gauge one obtains $\hat{P}\mathbf{A} - \mathbf{A}\hat{P} = 0$, and hence the equation (8) is rewritten as

$$\hat{H} = \frac{1}{2m} (\hat{P}^2 - 2q\mathbf{A}\hat{P} + q^2\mathbf{A}^2) + q\varphi. \quad (9)$$

Comparison between equations (7) and (9) shows that the potentials (\mathbf{A} , φ) participate in Quantum Mechanics exactly in the same way as in the classical Electrodynamics, namely, as vector and scalar functions of coordinates, respectively, rather than as differential operators.

References

- [1] Blokhintsev D I 1964 *Quantum Mechanics* D. Reidel, Dordrecht, Holland
- [2] Ershkovich A 2010 *Physics Today* **63**(4) 8
- [3] Landau L D, Lifshitz E M 2002 *Quantum Mechanics - Non-relativistic theory* Pergamon Press
- [4] Courant R, Hilbert D 1962 *Methods of mathematical physics* N-Y: Intersci. Publ. John Wiley & Sons **2**(9)
- [5] Aharonov Y, Bohm D 1959 *Phys. Rev.* **115** 485-91
- [6] Ershkovich A, Israelevich P 2013 *Aharonov-Bohm effect and classical Hamiltonian mechanics* Cornell Univ. Library <http://arXiv.org/abs/1105.0312>
- [7] Ershkovich A 2013 *Electromagnetic potentials and Aharonov-Bohm effect* Cornell Univ. Library <http://arXiv.org/abs/1209.1078>
- [8] Landau L D, Lifshitz E M 1975 *The classical theory of fields* Pergamon Press
- [9] Feinberg E L 1963 *Sov.Phys.Usp.* **5** 753-60 doi: 10.1070/PU1963v005n05ABEH003453, 1963

Author



Alexander Ershkovich

Current position: Professor Emeritus, Tel Aviv University

Scientific interest: Space Plasma Physics

Publications: more than 100

Experience: Solar wind interaction with planets and comets. Instabilities in Space Plasma. Helical waves in cometary plasma tails. Electric currents in geomagnetosphere.

Numerical simulation and hydrodynamic performance prediction for hydroplane longitudinal motion

Xiao Liang^{1*}, Wei Li², Weitong Fan¹, Guocheng Zhao¹

¹ Transportation Equipments and Ocean Engineering College, Dalian Maritime University, China

² College of Environmental Science and Engineering, Dalian Maritime University, China

Received 1 February 2014, www.tsi.lv

Abstract

Aiming at hydrodynamic performance prediction for hydroplane longitudinal motion, numerical simulation for a hydroplane motion was carried out by using VOF and RNG $k-\epsilon$ model and solving Navier-Stokes equation under FLUENT software platform. Evolution of ship resistance was obtained as the velocity change, and flow field situation and dynamic pressure variation of hydroplane hull bottom were reflected intuitively. By comparing and analysing the results among numerical simulation calculation and ship model experiments and theoretical estimation, it was verified that hydrodynamic performance prediction for hydroplane longitudinal motion based on numerical simulation calculation under FLUENT is feasible and precise enough.

Keywords: numerical simulation; hydrodynamic performance prediction; hydroplane; FLUENT

1 Introduction

A hydroplane is a high-speed craft, which depends on fluid dynamic pressure generated on craft body during traveling to support most of the body weight. Hydroplane is widely used for such good hydrodynamic performances as speedability, maneuverability, etc. when traveling at a high speed. The research on this ship type has also become a key research field for international scholars [1, 2]. However, the research on hydrodynamic performance prediction for hydroplane motion is difficult, which is usually carried out through ship model tests and theoretical approximate formula estimations [3, 4].

In 2003, Davidson Laboratory obtained the influence law of various main elements on the motion in waves through a series of sea keeping tests to prismatic hydroplane model in regular head waves and irregular waves [5]. A. Rosen and K. Garne from Sweden Royal Institute of Technology made a research on the hydrodynamics and dynamic pressure of V-type ships during traveling at a high speed in still water through ship model tests [6]. The Ship and Ocean Engineering Dept. of UK University of Glasgow used CFD and ship model testing method to make a research comparison on forces on water surface speedboat in still water [7]. Dong Wencai, et. al. from China Naval Engineering University made a basic assumption for the longitudinal motion of speedboats in head waves, established a basic equation for longitudinal motion considering the influence of dynamic elevating force and proposed a new method of longitudinal motion forecasting—hydroplaning method. This method was verified to forecast the longitudinal

motions in head waves at medium and high navigational speeds [8]. In recent years, with the rapid development of computer technology, numerical simulation has also had a great development with various pieces of CFD software being more and more powerful in function and being increasingly higher in calculation accuracy. Using numerical simulation calculation method to simulate hydroplane motion condition and study hydrodynamic performance prediction has become a solution with a significant meaning.

In this paper, VOF method is used on FLUENT platform and RNG $k-\epsilon$ model is combined to carry out numerical simulation calculation for hydroplane longitudinal motion by solving Navier-Stokes equation. Thereby the law of navigational resistance of hydroplane varying with the navigational speed is obtained and a comparison is made between the calculation results with FLUENT and the values of ship model test and theoretical estimation. The pressure variations on hydroplane hull bottom and the variations of flow field around hydroplane are investigated.

2 Control Equation and Numerical Calculation Method

For incompressible viscous flow under a typical Cartesian coordinate system, the influence of density impulsion is neglected and the time mean continuity equation, Reynolds equation in the form of tensor can be written as:

* Corresponding author - Tel: +86-0411-84727985; fax: +86-0411-84725960; E-mail: liangxiao19801012@126.com

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = 0, \quad (1)$$

$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \left(\mu \frac{\partial u_i}{\partial x_j} - \overline{\rho u_i u_j} \right) + S_i, \quad (2)$$

where $i, j = 1, 2, 3$; ρ is the fluid density; μ is the dynamic viscosity coefficient; u_i, u_j are the time mean values of speed components; u'_i, u'_j are the impulse values of speed components; p is the time mean value of pressure; S_i is general source item of momentum equation. The overline “—” means the time mean of physical quantities.

In RNG $k-\varepsilon$ model, the influence of small dimension is embodied in large dimension motion and corrected viscosity item to remove small dimension motion from the control equation systematically. The k equation and ε equation are obtained:

$$\frac{\partial(\rho k)}{\partial t} + \frac{\partial(\rho k u_i)}{\partial x_i} = \frac{\partial}{\partial x_j} \left(\alpha_k \mu_{eff} \frac{\partial k}{\partial x_j} \right) + G_k + \rho \varepsilon, \quad (3)$$

$$\frac{\partial(\rho \varepsilon)}{\partial t} + \frac{\partial(\rho \varepsilon u_i)}{\partial x_i} = \frac{\partial}{\partial x_j} \left(\alpha_\varepsilon \mu_{eff} \frac{\partial \varepsilon}{\partial x_j} \right) + C_{1\varepsilon}^* \frac{\varepsilon}{k} G_k - C_{2\varepsilon} \rho \frac{\varepsilon^2}{k}, \quad (4)$$

where k is the turbulence kinetic energy; ε is the turbulence kinetic energy loss rate.

$$\mu_{eff} = \mu + \mu_t, \quad \mu_t = \rho C_\mu \frac{k^2}{\varepsilon}, \quad C_\mu = 0.0845,$$

$$\alpha_k = \alpha_\varepsilon = 1.39, \quad C_{1\varepsilon}^* = C_{1\varepsilon} - \frac{\eta(1 - \eta/\eta_0)}{1 + \beta\eta^3}, \quad C_{1\varepsilon} = 1.42,$$

$$C_{2\varepsilon} = 1.68, \quad \eta = (2E_{ij} \cdot E_{ij})^{1/2} \frac{k}{\varepsilon}, \quad E_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right),$$

$$\eta_0 = 4.377, \quad \beta = 0.012$$

In comparison with standard $k-\varepsilon$ model, RNG $k-\varepsilon$ model consider the rotation in average flow and swirling flow by correcting the turbulence kinetic energy viscosity. An item was added in ε equation to reflect the time mean strain of main flow E_{ij} . The item generated in RNG $k-\varepsilon$ model is not only related to the flow condition, but also the function of space coordinates in the same question. As a result, flows with a high strain and highly bent flow lines can be treated better. As RNG $k-\varepsilon$ model is effective for sufficiently developed turbulence, and flows near walls has a low Re number where the turbulence is not sufficiently developed, with the influence of turbulence impulsion being not so high as that of molecular viscosity, no calculation can be made with that model in this area and a special means of

treatment must be used. Here the wall function method is used to treat the flows near walls.

For free surface treatment questions, VOF method is selected in this paper [9]. VOF constructs and tracks a free surface according to the function F of volume taken by a fluid in grid cells at different times. For a space area containing a fluid of both gaseous and liquid phases, scalar function f is defined, with f value being 1 in case of liquid space points existing and being 0 for other points not occupied by liquid. Integrate f value on grid cells and divide this integration value by cell volume to obtain the average value of f for cells, that is, cell volume taken by liquid in grid cells, which is defined as F . If $F = 1$ in a grid cell at some moment, it means that the grid cell is totally occupied by the fluid of a designated phase, being a fluid cell. If $F = 0$, that cell is totally occupied by the fluid of another phase. When $0 < F < 1$, the cell is a boundary cell containing materials of both phases. The equation that function F satisfies:

$$\frac{\partial F}{\partial t} + u \frac{\partial F}{\partial x} + v \frac{\partial F}{\partial y} + w \frac{\partial F}{\partial z} = 0. \quad (5)$$

3 Establishment and Setup of Calculation Model

The study object is a model “USV-3” used in hydroplane towing test. The ship model is a wooden model with a ground smooth surface sprayed with paint and the hull is naked. Through inspection, it is found to meet the tolerance standard in “Hydroplane Model Resistance Testing Method”. The model displacement $\Delta = 66\text{kg}$, the scale ratio $\lambda = 1:4.58$, the model length $L_{OA} = 2.4\text{m}$ and the molded breadth $B = 0.735\text{m}$.



FIGURE 1 Hydroplane model in towing test

As the hydroplane molded lines are complicated and mostly three-dimensional curves, it is difficult to generate a curved surface directly in the pre-treatment software GAMBIT, Maxsurf software is used first to generate three-dimensional curves with surfaces and bodies being generated in GAMBIT and a calculation model for hydroplane is eventually obtained (as shown in figure 2).

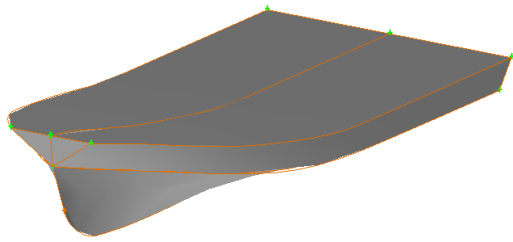


FIGURE 2 Calculation model of hydroplane

After a solid model is established in GAMBIT, adequate control field is selected according to the dimensions. According to the needs of flow field simulation calculation and by making reference to related literatures and experience from it, the control field is a rectangular body. The following scheme is used to set up the scope of calculation control field and the position of the ship model in the control field: the front end of the control field is at a location of $2 L_{OA}$ from the bow, and the back end is at a location of $3.75 L_{OA}$ from the stern, the uppermost boundary is at a location of $0.75 L_{OA}$ above the keel and the lowermost boundary is at a location of $2 L_{OA}$ below the keel and both sides are at the locations of $2 L_{OA}$ from the mid-section on the right and left. As the navigational speed of the hydroplane is high, there will be a long wake. Therefore, in order to capture the wake well, the control body is extended to make the back end locate at $5 L_{OA}$ from the stern. The solid ship model is 2.4m long, from which the length, width and height of the calculation control field obtained are 19.2 m, 9.6 m and 6.6 m respectively. As the model is symmetrical in relation to the longitudinal mid-section, half can be taken for the purpose of calculation. To simplify the model establishment process, the coordinate origin is set up at the lowest point of the stern, the direction of axis x pointing at the bow is positive, the direction of axis y pointing at the starboard is positive and the up direction of axis z is positive. The ship model position in the control field and the condition of control body is shown in figure 3.

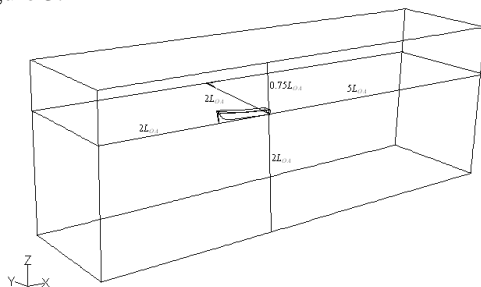


FIGURE 3 Ship model position of in the control field

Grid division is the most difficult part in establishing the mathematical model. How well the grid cells are divided not only determines if a correct solution can be obtained, but also determines the length of time to get the solution. As the grids divided in calculation vary greatly in volume from the location near the craft to the location far away from the craft, both non-structured and

structured grids are used in this paper. In consideration of computer calculation performance limitations and that the grids divided must change continuously in volume and there should not be too many grids and basic and key information on the flow field must be reflected, the control field is divided into several sub-fields in this paper, with non-structured grids being used near the craft and structured grids being used in a far location; the continuity of the grids between the sub-fields are controlled through the surface grids on the intersecting surfaces. Thus not only the need for calculation accuracy is ensured, but also the number of grids is reduced substantially and the calculation time is saved.

With regard to the locations containing a lot of small surfaces on the surface of craft body, those parts on the surface of the craft body with a small curvature variation are synthesized into larger surfaces first to reduce the difficulty in grid division. As the surface of the craft body is mostly composed of three-dimensional curved surfaces, it is difficult to generate structured grids near it and therefore surface grids are divided first on the surface of the craft body in which multiple verifications are necessary to find a suitable grid dimension. To facilitate body grid division near the craft body, we choose to use triangular surface grids on the surface of the craft body, and use tetrahedron non-structured grids near the craft body. To use quadrilateral grids on the surfaces of the calculation field can generate hexahedron structured grids in most areas in the control field except the locations near the craft body. The ratio between structured and non-structured grids will have a great influence on the calculation speed and quality. Generally, the use of non-structured grids should be reduced as possible as reasonably practical.

The areas near the craft body are important areas for the calculation study, where the density and quality of grids have a direct influence on the calculation results. In this paper, the method of increasing local grid density is used to refine grids and heighten the calculation accuracy. Grids of less density are used in areas farther away from the craft body and attention is paid to maintaining the grid density transition continuity. In dividing grids, the hexahedron structured grids in the far craft fields around are divided first and the tetrahedron non-structured grids near the craft are divided at last. This is favourable to grid transition between the sub-fields in the calculation field. The grid division in the control field is shown in figure 4.

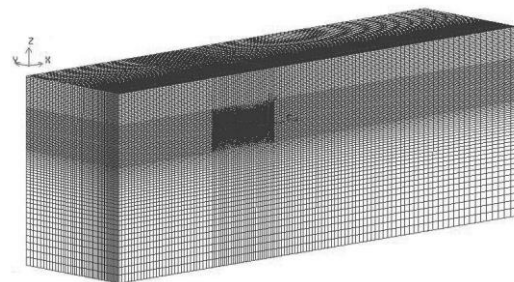


FIGURE 4 Grid division in calculation control field

4 Control Equation and Numerical Calculation Method

Numerical simulation calculation of hydroplane longitudinal motion is done on the FLUENT platform. Numerical simulations are made to the flow fields when a hydroplane makes even-speed longitudinal motion at five different navigational speeds, and a calculation study is made on the resistance. It is necessary to note that the corresponding trim angles and drafts of hydroplane at different navigational speeds can be obtained through ship model testing and theoretical analysis and are embodied in both GAMBIT modelling and FLUENT numerical calculation, which are not described in detail herein.

As shown in table 1 are the resistance value R_d and resistance coefficient C_d calculated by FLUENT and the comparison with ship model test values and the values estimated by applying empirical formulas. We can see that the resistance value R_d increases with navigational speed V or Froude number Fr_v , corresponding with the actual situation. The corresponding resistances calculated by FLUENT and the estimated resistance values at different speeds have no much difference, with the error between the corresponding calculated value and estimated value at 5 m/s being the largest as 3.56%.

TABLE 1 Resistance calculation comparison

Speed v [m/s]	5	6	7	8	9
Fr	2.51	3.01	3.52	4.02	4.52
FLUENT value F_R (N)	91.44	101.66	113.57	124.45	149.39
Resistance coefficient C_d	0.0084	0.0079	0.0053	0.0048	0.0046
Empirical estimation value (N)	102.31	104.06	111.28	123.96	141.25
Ship model test value (N)	94.82	100.40	111.32	126.86	147.06
Error in relation to test values %	-3.56	1.25	2.02	-1.9	1.59

Figure is the curve for FLUENT calculated values, test values, empirically estimated value varying with speeds. From the changing trends of the curves, they meet basically the law of resistances varying with speeds for hydroplane in a hydroplaning condition. Therefore, it is practical to apply FLUENT software to predict the resistance performances of a hydroplane in longitudinal motion and the accuracy is very high.

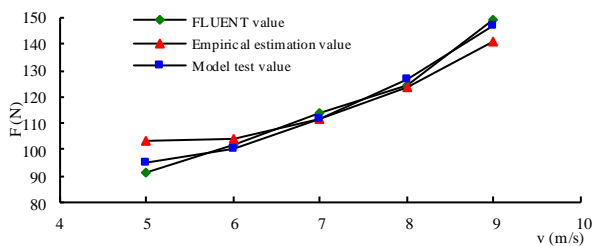


FIGURE 5 Curves for resistance varying with speed

The following is an investigation of the flow field change in numerical simulation process for hydroplane longitudinal motion. Figure 6(a)~(d) are cloud charts of volume fraction for the two phases of air and water on water surface varying with speeds, from which we can see intuitively the wake field varies with speed.

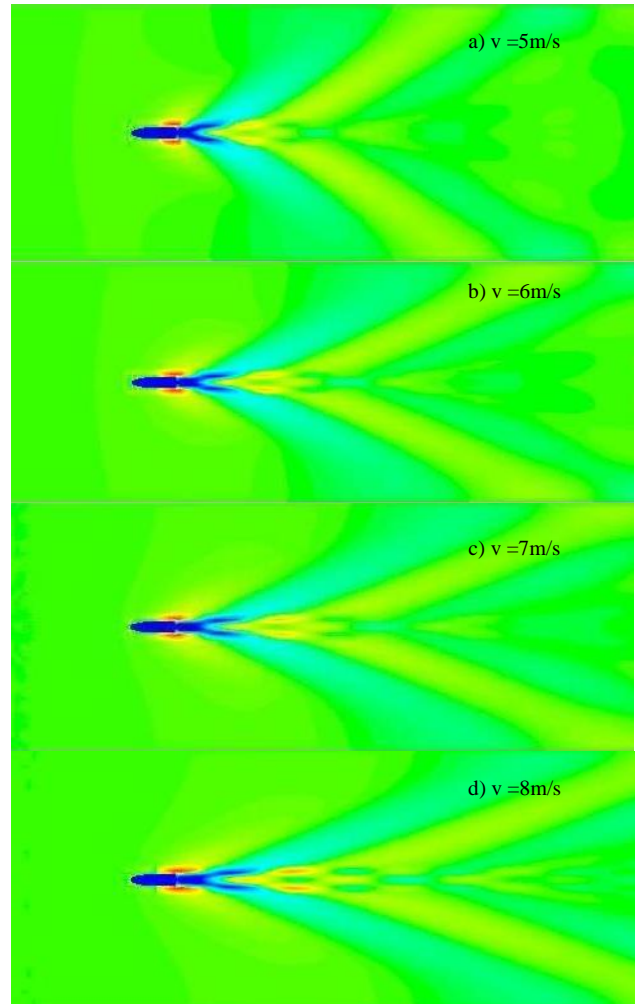


FIGURE 6 Cloud charts of volume fraction for the two phases of air and water on water surface varying with speed

Figure 7(a) ~ (d) are cloud charts of dynamic pressure on the hull bottom of hydroplane. As the navigational speed increases, the dynamic pressure on the hydroplane hull bottom increases. The dynamic pressure is a main reason for hydroplane to generate elevating force. Therefore, it indicates that the elevating force of the hydroplane increases with the navigational speed and this corresponds with the actual situation.

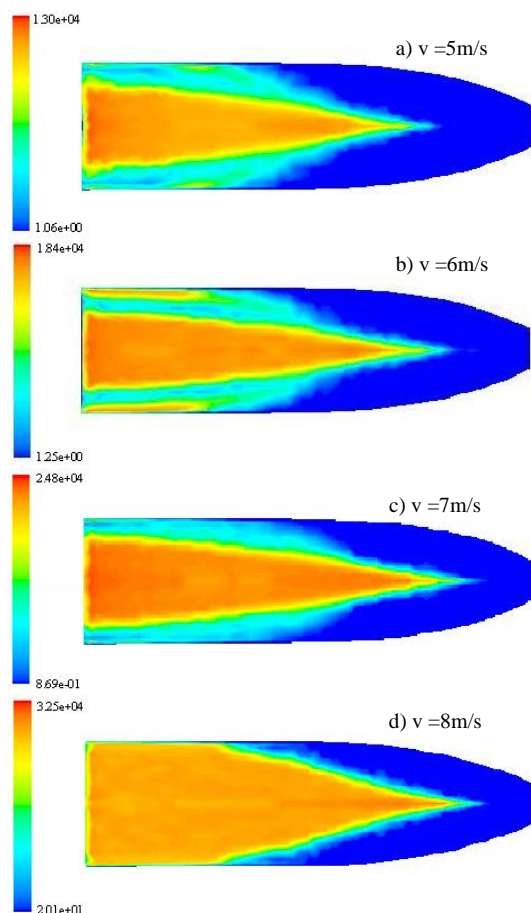


FIGURE 7 Cloud charts of dynamic pressure on hydroplane hull bottom varying with speed

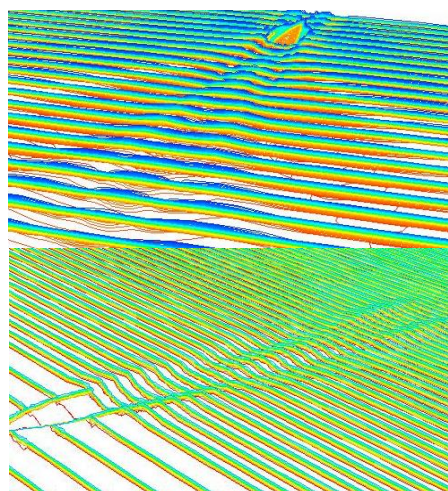


FIGURE 8 Wake flow of hydroplane motion

Figure 8 shows the change of wake flow during hydroplane longitudinal motion. From the simulation of flow field, the method can simulate wake flow of hydroplane motion well.

4 Conclusions

Aiming at hydrodynamic performance prediction for hydroplane longitudinal motion, numerical simulation for a hydroplane motion was carried out on FLUENT software platform. The law that the navigational resistance of hydroplane varies with the navigational speed is obtained. The variation of the flow field around hydroplane and the variation of the pressure on the hydroplane hull bottom are truly reflected. By comparing FLUENT calculated results and ship model test values and theoretically estimated values, it is proved that it is feasible and highly accurate to simulate water surface hydroplane motion and study the hydrodynamic performances on FLUENT platform.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 51209025, 51379026) and Fundamental Research Funds for the Central Universities of China (Grant No. 3132014067, 3132013328).

References

- [1] Wang B. 2008 *Marine Equipment/Materials & Marketing* 2 4-17
- [2] Wu Y S, Ni Q J, Ge W Z. 2008 *Journal of Ship Mechanics* 12(6) 1014-31
- [3] Peng G W 2003 *Journal of Wuhan Institute of Shipbuilding Technology* 3 32-35
- [4] Xu Y L, You D D. 2009 *Ship & Boat* 20(4) 4-7
- [5] Y.Ikeyda, T.Katayama 2000 *Physical & Engineering Sciences, Philosophical Transactions of The Royal Society (A)* 358(1771) 1905-15
- [6] A Rosen, K Garne 2004 *Transaction of the Royal Institute of Naval Architects* 299-308
- [7] Tveitnes T, Fairlie-Clarke A C, Varyani K S 2008 *Ocean Engineering* 35(14) 1463-78
- [8] Dong W C, Liu Z H, Wu X G, Yue G Q 2007 *Journal of Ship Mechanics* 11(1) 55-61
- [9] Wang F J 2004 *The Analysis of Computational Fluid Dynamics-Principles and Applications of CFD Software* Tsinghua University Press: Beijing (in Chinese)

Authors	
	<p>Xiao Liang, born in 1980, China</p> <p>Current position, grades: Dr. Liang is the senior member of International Association of Computer Science and Information Technology, and the member of Royal Institute of Naval Architect</p> <p>University studies: doctor degree in Design and Construction of Naval Architecture and Ocean Structure, Harbin Engineering University, Harbin, China, 2009</p> <p>Scientific interest: major field of study is hydrodynamics and underwater vehicles. One of his current jobs is Reliability Analysis of AUV Control System cooperating with State Key Laboratory of AUV, China. His current research interests include modelling, simulation and intelligent control of the AUV with fins</p> <p>Publications: more than 20 papers in journals and conferences</p>
	<p>Wei Li, born in 1980, China</p> <p>Current position, grades: One of her current jobs is Forecast and Emergency Response of Ship Oil Spill.</p> <p>University studies: received doctor degree in Environment Science and Technology, Harbin Institute of Technology, Harbin, China, 2008.</p> <p>Scientific interest: major field of study is hydrodynamics and numerical analysis. Her current research interests include modelling and simulation.</p> <p>Publications: more than 20 papers in journals and conferences.</p>
	<p>Weitong Fan, born in 1989, China</p> <p>Current position, grades: Now he is the postgraduate of Dalian Maritime University</p> <p>University studies: received bachelor degree in Naval Architecture and Ocean Engineering, Dalian Maritime University, Dalian, China, 2009.</p> <p>Scientific interest: major field of study is hydrodynamic performance of the ship.</p>
	<p>Guocheng Zhao, born in 1992, China</p> <p>Current position, grades: He majors in Naval Architecture and Ocean Engineering in Dalian Maritime University as a junior student. He is the squadron commissary in charge of studies. He has finished two national innovation projects, two provincial innovation projects during the undergraduate and he is elected as Dalian Merit Student and Excellent Student Cadre for his active performance in study and work.</p>

Research on adaptive AUV tracking control system based on least squares support vector machine

Song Xiao-ru^{1*}, Chen Hua², Song Bao-wei³, Wu Jie¹

¹ School of Electronic and Information Engineering, Xi'an Technological University, 710032, Xi'an, P.R.China

² School of Mechatronic Engineering, Xi'an Technological University, 710032, Xi'an, P.R.China

³ School of Marine Engineering, Northwestern Polytechnical University, 710072, Xi'an, P.R.China

Received 1 February 2014, www.tsi.lv

Abstract

Aimed at the nonlinear, uncertainties and impreciseness in AUV tracking control, an adaptive LSSVM control is highlighted. It is including the approximate dynamic inversion control law, linear controller and the robust adaptive LSSVM controller. The key is designed that adaptive LSSVM controller, which is to decrease or offset the uncertain dynamic inverse model errors. The online adjustment LSSVM parameters rules are deduced by the Lyapunov stability theorem. So the closed-loop tracking control system's stability and asymptotical convergence of tracking error can be guaranteed. It can be seen that the tracking errors converge and stay at a small neighbourhood of zero. At last taken a certain real AUV as an example, the numerical simulation results are given in the presence of the ocean current wave interference. They show the proposed method has good tracking performance and a certain robustness against modelling errors.

Keywords: AUV, LSSVM, tracking control, adaptive controller

1 Introduction

Since last decade AUV (Autonomous Underwater Vehicle, AUV) has gradually developed a new type of underwater vehicle with some type of autonomous navigational capability. It can be replaced manned vehicle, especially in dangerous tasks or difficult to reach the sea field. Thus, it is necessary to design an autonomous and precise guidance control system. The tracking control is essentially complete AUV motion control. The main task is accurately tracking a given path considering the AUV actual position and the state of motion. It is a very important link for AUV navigation. But the AUV has strong nonlinear uncertain parameters. And it is difficult to establish precise mathematical model. It is a more difficult problem to robust adaptive control.

Considered the relative merits of previous models, an adaptive LSSVM tracking control model for AUV is key put forward in the study. There is no dimension disaster and local minimum problems based on empirical risk minimization. It can solve the nonlinear uncertainty of error approximation problem during the AUV tracking control. And it has the high learning speed, good generalization and fine robustness. It can effectively prevent the over fitting phenomenon. The LSSVM provides a good effective means for nonlinear system identification and modelling. Therefore, the AUV tracking control is designed by the adaptive LSSVM

controller in the study.

The rest of the paper is organized as follows. The problem formulation of the AUV tracking control system is presented in the section 2. It is including the velocity and the heading subsystem. The next section is the key. The section 3 shows the robust adaptive controller based on the LSSVM. It consists of the approximate dynamic inversion control law, linear controller and the robust adaptive LSSVM controller. The simulation study of the adaptive LSSVM tracking controller is in the section 4. Firstly the AUV tracking control simulation model is given. Then the simulation results of the AUV tracking control are provided to demonstrate the robust adaptive LSSVM controller. At last, conclusions are drawn.

2 Problem Formulation of the AUV Tracking Control System

A complete six degree-of-freedom mathematical model for AUV can be found in reference [5]. Assume that the coupling was neglected the coupling between longitudinal and lateral movement. The uncertain AUV tracking equations can be simplified as described below.

$$\begin{cases} \dot{x}_v = x_a \\ \dot{x}_a = \bar{f}(x_v, x_a, u_1) + \tilde{f}(x_v, x_a, u_1), \\ y_1 = Cx = x_v \end{cases} \quad (1)$$

* Corresponding author. Tel: +86-029-86173097; fax: +86-029-86173097; E-mail: masha0422@163.com

$$\begin{cases} \dot{x}_v = x_r \\ \dot{x}_r = \bar{f}(x_v, x_r, u_2) + \tilde{f}(x_v, x_r, u_2) \\ y_2 = Cx = x_v \end{cases} \quad (2)$$

$x_1 = (x_v, x_v)^T$, $x_2 = (x_a, x_r)^T$, $u = (u_1, u_2)^T$, $y = (y_1, y_2)^T$, where $x_1, x_2 \in R^n$, x_1 is the control output. x_v and x_r is the AUV velocity and the yaw output, respectively. x_2 is the derivative of the AUV speed and yaw. Namely, they are acceleration and angular rate. The control put is $u \in R^m$. The u_1 and u_2 are the rudder control input. The output is $y \in R^n$. C is the appropriate dimensional matrix. \bar{f} is the nominal value. \tilde{f} is uncertainties.

Note $x = (x_1^T, x_2^T)^T$, $f = \bar{f} + \tilde{f}$.

The asymptotic stability reference model is:

$$\begin{cases} \dot{x}_{m1} = x_{m2} \\ \dot{x}_{m2} = -K_{m1}x_{m1} - K_{m2}x_{m2} + K_{m1}r \\ y = C_m x_m = x_{m1} \end{cases} \quad (3)$$

where $x_{m1}, x_{m2} \in R^n$, $x_m = (x_{m1}^T, x_{m2}^T)^T$, $r \in R^n$ is the input of the reference model. $y_m \in R^n$ is the output of the reference model. C_m and $K_{mi}(i=1,2)$ are the optimal dimensional matrixes. The control goal is to design robust adaptive control law u makes the closed loop system stable. And the output y can track the any given reference signal y_m .

3 Design of Robust Adaptive Controller

The robust adaptive controller is including the approximate dynamic inversion control law, linear controller and the robust adaptive LSSVM. The first one is approximate dynamic inversion control law. It is to decrease or offset the nonlinear and time-variant characteristic. The online LSSVM is to minish or counteract uncertain dynamic inverse model errors, which caused by model inaccuracy and uncertainty. Linear control law makes the system response has the desired quality. The principle block diagram of robust adaptive controller is as the following Figure.1.

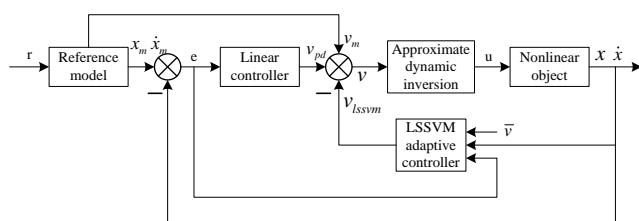


FIGURE 1 The principle diagram of robust adaptive controller

3.1 APPROXIMATE DYNAMIC INVERSE CONTROL

Suppose

$$\bar{f}(x, u) = v, \quad (4)$$

where v is the pseudo-control quantity. The approximate dynamic inverse control law of system (1) are obtained by the Formula (4) of inverse.

$$u = \bar{f}^{-1}(x, v). \quad (5)$$

It must be pointed out that appropriate modify $\bar{f}(x, u)$ to $\hat{f}(x, u)$ to ensure the existence of approximate dynamic inversion control law when there nonexistence the inverse of $\bar{f}^{-1}(x, u)$. Such the introduction of the model inversion errors are compensated by the LSSVM. The approximate dynamic inversion errors often cannot be completely offset the nonlinear because of inaccurate, uncertain factors, interference and even system failure. Therefore, there exists dynamic inversion error.

Note the inverse error is

$$\tilde{f} = f(x, u) - \bar{f}(x, u), \quad (6)$$

$$\ddot{x}_1 = v + \tilde{f}. \quad (7)$$

3.2 DESIGN OF LINEAR PD CONTROLLER

Set pseudo-control quantity

$$v = v_{pd} + v_m + v_{lssvm}, \quad (8)$$

where

$$\begin{cases} v_{pd} = K_p(x_{m1} - x_1) + K_D(\dot{x}_{m1} - \dot{x}_1) \\ v_m = \ddot{x}_{m1} \end{cases} \quad (9)$$

where, K_p and K_D are the appropriate dimensional proportional and differential parameters matrix. v_{pd} is the output of the linear controller. It is the proportional derivative control. v_m is the signal from the reference model. v_{lssvm} is the LSSVM adaptive output signal. It is used to offset the inverse error \tilde{f} . x_{m1} is the status signals of the reference model.

The Eq. (8) and Eq. (9) are substituted into the formula (7). After the compilation, it can be tracking error dynamic characteristics. It is as follows.

$$\dot{e} = Ae + B(v_{lssvm} - \tilde{f}). \quad (10)$$

The system matrix A can be a Hurwitz matrix by selecting reasonable parameters K_p and K_D . If the K_p and K_D are diagonal matrix the Lyapunov equation is constructed as follows.

$$PA + A^T P = -I. \tag{11}$$

Positive definite symmetric solution is

$$P = \begin{bmatrix} \frac{K_D K_p^{-1} + K_p K_D^{-1}}{2} & \frac{K_p^{-1}}{2} \\ \frac{K_p^{-1}}{2} & \frac{K_D^{-1}}{2} (1 + K_p^{-1}) \end{bmatrix}. \tag{12}$$

Theorem 1: To the given asymptotically stable reference model (3), the state of the system can be completely track the reference model output when $K_p = K_{m1}$, $K_D = K_{m2}$ and there is no inverse error or the inverse error completely offset by using LSSVM.

Prove: it is gained between the Eq. (7) and Eq. (8).

$$\ddot{x}_1 = K_p x_{m1} - K_p x_1 + K_D \dot{x}_{m1} - K_D \dot{x}_1 + \ddot{x}_{m1}. \tag{13}$$

Let Eq. (3) into equation (13). It can be gained when $K_p = K_{m1}$ and $K_D = K_{m2}$.

$$\dot{x}_1 = x_2, \dot{x}_2 = -K_{m1} x_1 - K_{m2} x_2 + K_{m1} r.$$

Thus, it can be seen the state of the system response is the reference model state response.

3.3 THE DESIGN OF THE ADAPTIVE LSSVM CONTROLLER

3.3.1 LSSVM Model

The LSSVM is proposed by Suykens. It is used to solve classification and function estimation problems. The function estimation algorithm of LSSVM is deduced briefly as follows. Suppose the training sample is a n-dimensional vector, and N samples form the sample set D , $D = \{(x_i, y_i) | i=1,2,\dots,N\}$, $x_i \in R^n$, $y_i \in R$, where x_i is an input data, and y_i the corresponding output value. LSSVM can map the data to a high-dimensional feature space by using a nonlinear mapping ϕ . Then, the equation of linear regression is

$$y(x) = w^T \phi(x) + b, \tag{14}$$

where $w \in R^{n_h}$ is a weight vector, b is a constant bias.

The learning process can be translated into an optimization problem according to the principle of structural risk minimization.

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^N e_i^2, \text{ s.t.}$$

$$y_i = w^T \phi(x_i) + b + e_i, i=1, 2, \dots, N. \tag{15}$$

$\phi(x): R^n \text{ to } R^{n_h}$ is a mapping function in the kernel space, $e_i \in R$ is the error variance, J is the loss function, C is the adjustable regularized parameter.

A Lagrangian function can be introduced to solve the above optimization problem, which can transform the optimization with constraint to the optimization without constraint.

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i (w^T \phi(x_i) + b + e_i - y_i), \tag{16}$$

where $\alpha_i \in R$ is a Lagrangian multiplier. The partial derivatives of L with respect to w , b , e_i and α_i are found out according to Karush-Kuhn-Tucker (KKT) condition. The middle variables w and e_i are eliminated, and the solution of α_i and b can be obtained

$$\begin{bmatrix} 0 & \eta \\ s & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}, \tag{17}$$

where $\Omega \in R^{N \times N}$, $\Omega_{km} = \phi^T(x_k) \phi(x_m)$, $Y = [y_1, \dots, y_N]^T$, η is a N-dimensional column vector, namely, $\eta = [1 \dots N]$, $\alpha = [\alpha_1, \dots, \alpha_N]$, I is a confirmed matrix, S is a N dimensional row vector, namely, $s = [1 \dots N]^T$.

So the function estimation can be expressed as

$$y(x) = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b, \tag{18}$$

where $K(x_i, x_j)$ is the kernel function that meets Mercer requirements [9]. Gaussian radial basis function (RBF) is selected as the kernel function in this paper.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\delta^2}\right), \tag{19}$$

where $K(x_i, x_j)$ is the kernel function that meets Mercer requirements [9]. δ is the kernel factor, i.e. the width of the kernel function.

3.3.2 The structure of the adaptive LSSVM controller

The dynamic inversion error signal is approached by the LSSVM. The outputs of support vector machine are the pseudo-linear system control inputs. The errors are online eliminated. The structure drawing of the adaptive LSSVM is as followings figure. 2.

In figure 2 $\bar{x} = [x^T, \dot{x}^T, e^T, \dot{e}^T, \bar{v}^T]^T$ is the input vector. \bar{v} is a signal after the pseudo control amount V saturated nonlinear processed. The hidden layer nodes is $N+1$. Where, the first node is defined as the deviation of the hidden layer. Its value is 1 from the formula (18). $w_j(j=1,2,\dots,N,N+1)$ are the weights of the hidden layer and the output one, where $w_1 = b$, $w_j = a_{j-1}(j=2,3,\dots,N,N+1)$.

The input-output relation of LSSVM is

$$v_{lssvm} = W^T \beta, \tag{20}$$

where $W = [w_1, w_2, \dots, w_{N+1}]^T$, $\beta [1, K(x_1, \bar{x}), \dots, K(x_N, \bar{x})]$.

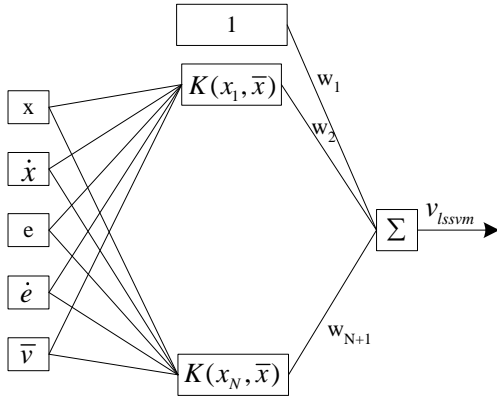


FIGURE 2 The structure drawings of the adaptive LSSVM

There exists the optimal weight vector W^* when reconstruction error $\varepsilon < 0$ for the continuous uncertain nonlinear inverse error function \tilde{f} and any given inverse error because LSSVM algorithm can be approached the continuous nonlinear function in arbitrary precision. It makes the following equation is established.

$$|W^{*T} \beta - \tilde{f}| \leq \varepsilon, \quad \varepsilon > 0. \tag{21}$$

The system tracking error dynamic equation (10) is re-writing the following formula:

$$\dot{e} = Ae + B\tilde{W}^T \beta + B(W^{*T} \beta - \tilde{f}), \tag{22}$$

where $\tilde{W} = W - W^*$. W^* is the optimal weight vector.

Theorem 1: For the AUV velocity formula (1) and heading (2) system, Least squares support vector machine dynamic inversion error compensated control structure is as Fig.1. The analytical expression of the control structure is the formula (8). The adaptive compensation term is the equation (20). The all signals in the closed-loop tracking control system remain bounded when the

law of weight adjustment is the formula (23). Where $\gamma > 0$ is the adaptive gain. P is the equation (12).

$$\begin{cases} \dot{W} = \tilde{W} = -\gamma e^T P B \beta, & \|e\| \geq e_0 \\ 0, & \|e\| \leq e_0 \end{cases}. \tag{23}$$

Prove: the following Lyapunov function is selected.

$$V = \begin{cases} \frac{1}{2} e^T P e + \frac{1}{2\gamma} \tilde{W}^T \tilde{W} & \|e\| \geq e_0 \\ E_0 + \frac{1}{2\gamma} \tilde{W}^T \tilde{W} & \|e\| \leq e_0 \end{cases}. \tag{24}$$

The E_0 in the equation (24) is satisfied

$$E_0 = \frac{1}{2} e^T P e, \quad \|e\| = e_0. \tag{25}$$

It can be ensure that V is continuous when the error is at the boundary of the dead zone.

There is equation (26) outside the dead zone $\|e\| \geq e_0$.

$$\dot{V} = \frac{1}{2} \dot{e}^T P e + \frac{1}{2} e^T P \dot{e} + \frac{1}{\gamma} \tilde{W}^T \dot{\tilde{W}}. \tag{26}$$

It can be gained by the formula (22) and (23).

$$\dot{V} = -\frac{1}{2} e^T I e + e^T P B [W^{*T} \beta + (W^{*T} \beta - \tilde{f})] + \frac{1}{\gamma} \tilde{W}^T \dot{\tilde{W}}. \tag{27}$$

By the formula (21)

$$\dot{V} \leq -\frac{1}{2} e^T I e + \varepsilon |e^T P B| + \tilde{W}^T \left(e^T P B \beta + \frac{1}{\gamma} \dot{\tilde{W}} \right). \tag{28}$$

The adaptive law can be got by using the formula (23).

$$\begin{aligned} \dot{V} &\leq \left(-\frac{1}{2} e^T I e + \varepsilon |e^T P B| \right) \\ &\leq \left(-\frac{1}{2} \|e\| \lambda_{\min}(I) \|e\| - 2\varepsilon \lambda_{\max}(P) \right), \end{aligned} \tag{29}$$

$$\begin{aligned} \therefore \|B\| &= 1 \\ \therefore \dot{V} &< 0 \end{aligned}$$

when

$$\|e\| > \frac{2\varepsilon \lambda_{\max}(P)}{\lambda_{\min}(I)} > 2\varepsilon \lambda_{\max}(P). \tag{30}$$

So if $\|e\| > e_0$, then $\dot{V} < 0$. If $\|e\| \leq e_0$, then $\dot{V} = 0$.

Because of $\dot{\tilde{W}} = 0$ in the formula (23).

In short V has the following properties.

- (1) The V is a continuous when $V \geq E_0 > 0$ and $\|e\| = e_0$;
- (2) $\dot{V} < 0$ when $\|e\| > e_0$; $\dot{V} = 0$ when $\|e\| \leq e_0$;
- (3) Assume $\|e(t_0)\|$ and $\|\tilde{W}(t_0)\|$ are bounded, then $V \geq \lambda_{\min}(P)\|e\|^2/2 + \|\tilde{W}\|^2/2$.

The properties (1)~(3) mean that $\|e\|$ and $\|\tilde{W}\|$ are uniformly bounded

When the $\varepsilon=0$ the dead zone definition in the equation (30) is zero. The formula (22) and (23) can be changed into the followings.

$$\dot{e} = Ae + B\tilde{W}^T \beta, \tag{31}$$

$$\dot{\tilde{W}} = -\gamma e^T P B \beta. \tag{32}$$

The above two equation equilibrium point is the origin $(e=0, \tilde{W}=0)$. The corresponding Lyapunov functions are taken as.

$$V = \frac{1}{2} e^T e + \frac{1}{2\gamma} \tilde{W}^T \tilde{W}, \tag{33}$$

$$\dot{V} = -\frac{1}{2} e^T e. \tag{34}$$

The equation state is consistent with stable is described between the equation (31) and (32) from the equation (33) and (34).

The \dot{e} is bounded from the equation (31). Because

$$\frac{1}{2} \int_{t_0}^{\infty} e^T e d\tau = -\int_{t_0}^{\infty} \dot{V} d\tau = V(t_0) - V(\infty) < \infty. \tag{35}$$

So the e is the square integrable. Because \dot{e} and e are bounded, and e square integrable the formula (36) is gained by the Barbalat theorem.

$$\lim_{t \rightarrow \infty} e(t) \rightarrow 0. \tag{36}$$

So the equation (22) is asymptotically stable by $\varepsilon = 0$.

4 Simulation Study of the Adaptive LSSVM Tracking Controller

To verify the robust adaptive controller's effectiveness, taking a real certain vehicle as objective, the tracking

control simulations are conducted. Assume there are 60° ocean current direction and two sea state wave forces in the simulation environment. The AUV tracks the circle path under the specified speed. Suppose the AUV track along a radius of 4.8m, and the track centre at (0,4.8) circular path. The middle uniform tracking speed is 1.4m/s. the acceleration is the 0.2m/s² in the initial and end times. Select 300 data as the training samples, the other 200 data as the test samples. The simulation results are described as follows.

The numerical results are shown in figures. 3-9. the many AUV curves are given. Such as the displacement and velocity, the displacement and heading, the changes sailing speed and heading, the tracking error in the x-direction and y-direction. At last the AUV tracking the circle path by the adaptive LSSVM controller. It can be seen that the tracking errors converge and stay at a small neighbourhood of zero.

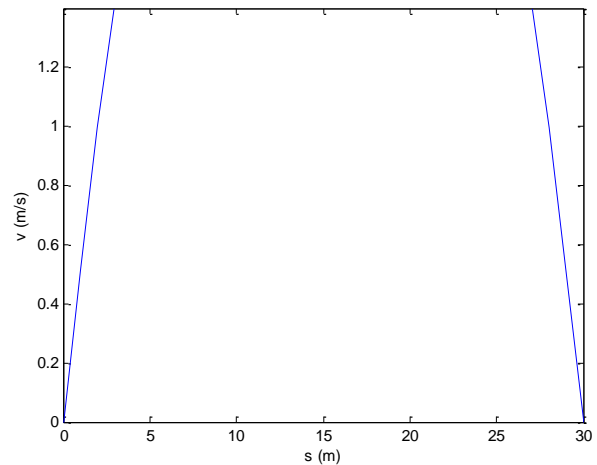


FIGURE 3 The change curve between the displacement and velocity

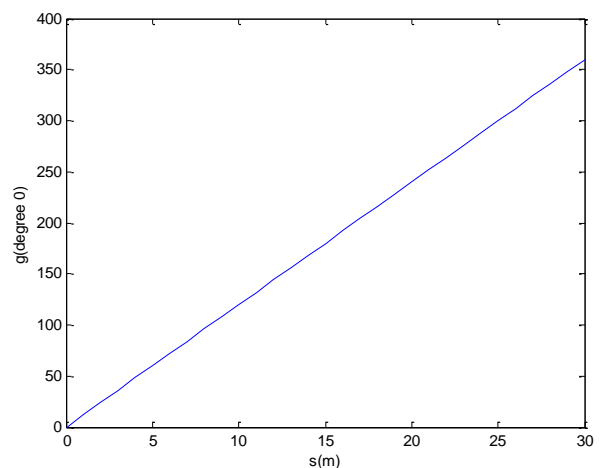


FIGURE 4 The change curve between the displacement and heading

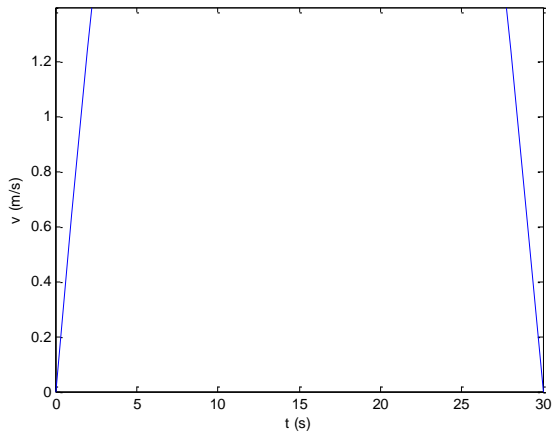


FIGURE 5 The changes sailing speed curve

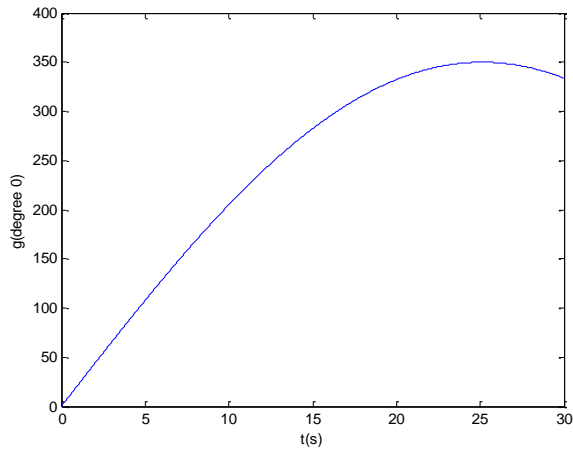


FIGURE 6 The AUV changes heading curve

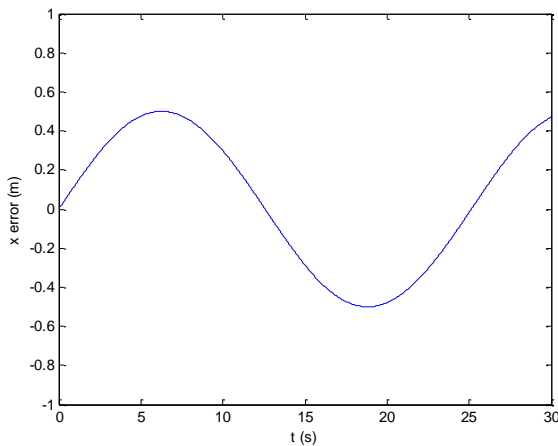


FIGURE 7 The AUV tracking error in the x-direction

The figure 3 shows the AUV has been achieved a desired speed 1.4m/s about less than 5 meters. And it meets the requirements before less than 5 meters distance to the end. The good effect is obvious. The change curve between the displacement and heading is showed in the figure 6. Generally, the yaw angle was linearly increasing trend with the voyage increase. The yaw angle of 360

degrees has been gained when the AUV about runs the full circle, namely the sail distance is about 30 meters. The preferably expectations are reached.

The change sailing speed curve is showed in the Figure 5. The specified speed 1.4m/s is reached with the 0.2m²/s acceleration in several seconds. And the velocity is reached zero within the a few seconds at the end of the voyage. The system has good rapidity and high accuracy. The Figure6 showed the AUV changes heading curve. The yaw angle is to the 360 degree at about 25s. Namely, the AUV can fully reach tracking the whole circle path. In ideal the all round is about 21s. The heading angle control effect is good considering the rise and fall time.

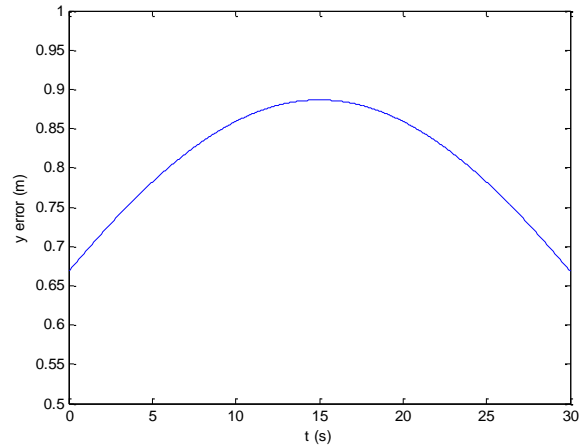


FIGURE 8 The tracking error in the y-direction

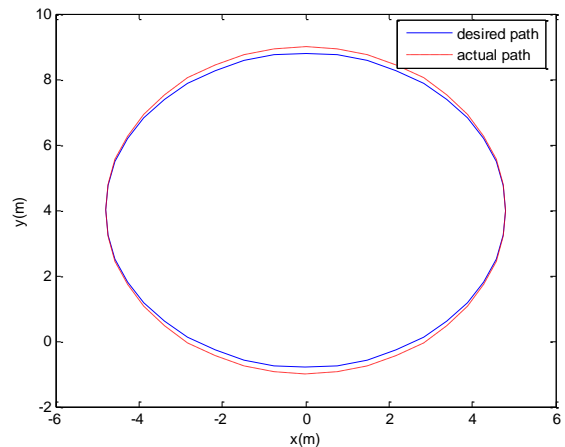


FIGURE 9 Tracking the circle path by the adaptive LSSVM controller

The AUV tracking error is displayed in the x-direction and y-direction from the figure 8 and 9, respectively. The errors do not exceed 1m. The error is very small. The control accuracy is quite satisfactory. At last, the AUV circular path tracking in the horizontal plane in the figure 10. The blue solid line is the ideal path. And the red dotted line is the corresponding actual path. It can be seen there is a certain degree of error between the two situations. However, the error is still in

the relatively small area. The overall tracking performance is quite satisfied.

5 Conclusions

In this study the problem of AUV trajectory-tracking has been investigated by the robust adaptive LSSVM controller. The design of robust adaptive controller is focused on. It is including the approximate dynamic inversion control law, linear controller and the robust adaptive LSSVM controller. The third part the robust adaptive LSSVM controller is the key built and designed. The LSSVM model can be approached a continuous nonlinear function in arbitrary precision by the adjusting the weight matrix. The online LSSVM is to minish or counteract the uncertain dynamic inverse model errors, which caused by model inaccuracy and uncertainty. It make the system output can be tracked the output of the reference model as small as possible errors. The online adjustment LSSVM parameters rules are deduced by the Lyapunov stability theorem. So the closed-loop tracking control system's stability and asymptotical convergence of tracking error can be guaranteed. It can be seen that the tracking errors converge and stay at a small neighbourhood of zero. At last taken a certain real AUV as an example, the numerical circle path simulation results are given in the presence of the ocean current wave interference. The numerical results have been showed that the proposed method has good tracking performance and certain robustness against modelling errors. The method have been improved the model's accuracy, reliability and strong anti-interfere capability.

Acknowledgments

The project is financially supported the Education Department of Shaanxi Provincial Government Research

Project and the Xi'an Science and Technology Bureau Project. The authors would like to thank the underwater vehicle Institute of Marine Engineering, Northwestern Polytechnical University. The parts of experiment have been done by their help and support. And the authors would like to express my gratitude to my friends who had helped me!

References

- [1] Fierro R, Lewis L 1995 Control of a Nonholonomie Mobile Robot Using Neural Networks. *IEEE International Conference on Intelligent Control*, Monterey, CA
- [2] Petresand C, Patron P 2005 Path Planning for unmanned underwater vehicles, in *Proceeding so the IJCAI 2005 Workshop on Planning and Learning in A Priori Unknown or Dynamic Domains* Edinburgh, UK, August 2005
- [3] Seung-Keon Lee, Kyoung-Ho Soh 2009 Modelling and controller design of manta-type unmanned underwater test vehicle *Journal of Mechanical Science and Technology* 23(4) 987-990.
- [4] Ahmad Bagheri, Jalal Javadi Moghaddam 2009 Simulation and tracking control based on neural-network strategy and sliding-mode control for underwater remotely operated vehicle *Neurocomputing* 72(7-9) 1934-1950
- [5] Fossen T I 1994 *Guidance and Control of Ocean Vehicles* John Wiley& Sons Inc, New York
- [6] Suykens J A K 2001 Nonlinear modelling and support vector machines[C] *Proceedings of the IEEE Conference on Instrumentation and Measurement Technology*. Piscataway, NJ, USA IEEE
- [7] Javadi-Moghaddam J 2010 An adaptive neuro-fuzzy sliding mode based genetic algorithm control system for under water remotely operated vehicle *Expert Systems with Applications* 37(1) 647-660
- [8] Hartman R, Hawkinson W, Sweeney K 2008 Tactical underwater navigation system position *IEEE Location and Navigation Symposium*
- [9] Bahador Khaleghi 2011 Multisensor data fusion: a review of the state-of-the-art *Information Fusion* 12(8)
- [10] BI Feng-yang, WEI Ying-jie 2012 A Cascade Approach for Global Trajectory Tracking Control of Underactuated AUVs *Ocean Engineering* 24(2) (In Chinese)

Authors	
	<p>Song Xiaoru, born in April 1979, Xi'an, Shaanxi, China</p> <p>Current position, grades: PhD in Control Science and Engineering in Northwestern Polytechnical University</p> <p>University studies: M.Sc. in Control Theory and Control Engineering from Xi'an University of science and technology</p> <p>Scientific interest: information fuse and the AUV intelligent navigation and control etc.</p> <p>Publications: She has (co-) authored more than 2 books and 20 papers, which have 9 papers, were indexed by EI</p>
	<p>Hua Chen, born in July, 1962, Xi'an, Shaanxi, China</p> <p>Current position, grades: professor in Xi'an Technological University</p> <p>University studies: M.Sc. in Mathematics and PhD in Information Sciences</p> <p>Scientific interest: Pattern Recognition, reliability control etc.</p>
	<p>Baowei Song, born in July, 1963, Huludao, Liaoning, China</p> <p>Current position, grades: full professor of informatics at Informatics Department, Northwestern Polytechnical University</p> <p>University studies: M.Sc. in Mathematics and PhD in from Northwestern Polytechnical University</p> <p>Scientific interest: information fusion, AUV reliability etc.</p>
	<p>Jie Wu, born in February, 1979, XianYang, Shaanxi, China</p> <p>University studies: B.E. degree in Control Science and Engineering from Central South University in 2001, and M.Eng. degree in Control Science and Engineering from Xi'an Technological University in 2005. In 2006, he studied at Chinese Key Laboratory for Manufacturing Systems Engineering at Xi'an Jiaotong University for Ph.D. degree.</p> <p>Scientific interest: optimization of complicated systems, robotics and automation, control, and computer vision</p>

A study on capacity of signalized intersection under snow-ice conditions based on classical model modification

Yaping Zhang¹, Chunxiao Liu^{1*}, Hong Zhang², Guozhu Cheng¹

¹ School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, 150090, P. R. China

² School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Received 1 February 2014, www.tsi.lv

Abstract

As the common climatic phenomena frequently occurring in northern China in winter, snow and ice weathers have made great influence to the capacity of signalized intersection. By starting from snow and ice conditions, this paper makes improved correction about the calculation of capacity of signalized intersection respectively based on HCM Method and Method of Stopping Line. On the basis of HCM Method, the author comprehensively considers factors influencing the capacity of signalized intersection under snow and ice conditions, and calibrates the integrated parameters influencing the capacity of signalized intersection under snow and ice conditions; grounded on Method of Stopping Line, the author then calibrates various parameters and makes classified study about each parameter separately according to snow weather and freezing condition, finally giving out parameter values under various states. With the help of the revised model, the paper calculates the capacity of signalized intersection and then makes the comparison between traffic capacities under various snow and ice conditions and that with capacity of signalized intersection under normal weather, hence with strong practical significance.

Keywords: Snow and ice conditions; Signalized intersection; capacity; Parameter modification; Comparative analysis

1 Introduction

With the progress of society, people have put forward higher requirements for traffic network, especially the service level of intersection and the adaptability to changes in sudden traffic problems. Under snow and ice conditions, the friction coefficient on pavement/road surface decreases and the running speed of vehicles also reduces, both of which not only lower the capacity of road and aggravate traffic congestion, but also especially easily form the chain crowding effect, thus causing local congestion in road network. Under this circumstance, it can effectively relieve problems of traffic jam and declining capacity under snow and ice weathers and thus displays very practical significance for lowering occurrences of traffic accidents under snow and ice conditions to study the traffic situation of signalized intersection in traffic network, grasp its real service level and to take various effective measures about traffic management and control as well as induction.

At present, domestic scholars have proposed various methods to calculate the capacity of signalized intersection under adverse environments and the most typical is the method of reducing coefficient put forward in American Highway Capacity Manual [1]. Based on previous studies, such scholars as Mr.Zhang regard the feeling of drivers as main evaluation criterion under snow-ice conditions and then they integrate this with such factors as road operating capacity, traffic security

coefficient and outside wealth condition, where delay, control, weather and etc. are main factors influencing the capacity of signalized intersection under snow and ice weathers [2]. Lida and other scholars design the model with relative intensity quantization process to visualize feelings of drivers, and with the assistance of computer programs, it can vividly describe the process of drivers forming judgments in signalized intersection and their psychological anxiety changes, hence determining the service level of signalized intersection with this as the evaluation criterion [3]. With grey theory system as research basis, other scholars like Li establish the evaluation model with five important indicators like stopping delay and queuing length, and then evaluate the capacity of signalized intersection and its running state based on this model [4]. Such scholars as Peeheux, Zhang and Flannery further study comprehensive evaluation indices for motor vehicle service level in signalized intersection and design the evaluation index system based on this [5-7]. After collecting data about friction coefficient in different pavements with snow and ice and analysing the braking response time and braking length for drivers to start, stop vehicle on road with snow and ice and to make a turn at intersection, Cheng Guozhu and other scholars finally take the sedan car as example to give out the suggestive value for maximum speed ensuring safe traffic on pavement covered with snow and ice, thus calculating the capacity of signalized intersection under snow and ice conditions based on this

* Corresponding author - E-mail: 877289604@qq.com

[8]. To sum up, domestic studies about the capacity of signalized intersection under snow and ice conditions have not yet been mature, which is mainly because the psychological feelings and driving fatigue of drivers are difficult for quantitative research [9-10].

2 Modelling method

The capacity of intersection refers to the sum of vehicles that can pass each in-lane at intersection in unit time, and the model for capacity of intersection under snow and ice conditions is based on the capacity of signalized intersection under normal weather. After revising its value, the paper finally obtains the capacity of signalized intersection under snow and ice weathers, and the author then makes further research about this with HCM Method and Method of Stopping Line.

2.1 HCM METHOD

After analysing and systemizing lots of data and referring to the achievements of previous scholars, the paper gives the model to calculate the saturation flow rate of intersection lane under snow and ice conditions on the ground of HCM Method, which is shown in Formula 1 as follows:

$$S_{snow-ice} = f_{snow-ice1} \times S_0 \cdot N \cdot f_w \cdot f_{HV} \cdot f_g \cdot f_p \cdot f_{bb} \cdot f_a \cdot f_{LU} \cdot f_{LT} \cdot f_{RT} \cdot f_{Lpb} \cdot f_{Rpb} \cdot S_0 \quad (1)$$

In the formula, $f_{snow-ice1}$ is the correction coefficient of saturation flow rate under snow and ice conditions.

According to Formula 1, when it comes to calculating the saturation flow rate of each lane, firstly, the paper needs to select the ideal theoretical saturation flow rate S_0 . Generally speaking, each straight-going lane selects 1,650pcu/h while each left-turning and right-turning lane both select 1,550pcu/h. Then, the author makes various corrections about the value according to the influencing factors.

After determining the saturation flow rate of lane, the author sums the saturation flow rate of each single in-lane so as to obtain the capacity of intersection. Under snow and ice conditions, the calculation formula for capacity of each land group $C_{snow-icei}$ is shown in Formula 2 as follows.

$$C_{snow-icei} = S_{snow-icei} \times \lambda_i \quad (2)$$

In the formula, λ_i is the green-time-rate (the effective green time/signal cycle length);

$S_{snow-icei}$ is the saturation flow rate of lane or lane group i.

Then the capacity of the whole intersection is shown in the following Formula 3.

$$C_{snow-ice} = \sum_i C_{snow-icei} \quad (3)$$

In conclusion, in order to determine the capacity of signalized intersection under snow and ice conditions, the paper also needs to make clear the correction coefficient on snow and ice pavement $f_{snow-ice1}$ for the coefficient is influenced by several factors and it is also complicated. As a result, the author will make detailed research about the correction coefficient of saturation flow rate under snow and ice weathers in the next section.

2.2 METHOD OF STOPPING LINE

According to the road traffic flow situation in China, the author also needs to regard the stop line as control section when calculating the capacity of intersection controlled by computer signal; in other words, whether it involves with left turn, right turn or straight going, as long as the vehicle passes the stop line in effective green time, it can be deemed as entering and passing the intersection. Under snow and ice conditions, the method to calculate the capacity of intersection based on Method of Stopping Line is listed as follows:

(1) Capacity of one straight-going lane

$$C_s = f_{snow-ice2} \frac{3600}{T} \left(\frac{t_g - t_0}{t_i} + 1 \right) \quad (4)$$

In the formula, T is the cycle of signal light, which takes value according to the actual signal cycle.

C_s is the design capacity of one straight-going lane, (with the unit of) veh/h.

t_g is the green time in a signal cycle, s.

t_0 is the time for the first vehicle to start and pass the stop line after the green signal light is on, s, which averagely takes the value of 2.3s under normal weather.

t_i is the average time headway for straight-going vehicles to pass the stop line, s.

(2) Capacity of special right-turning lane

$$C_r = f_{snow-ice2} \frac{3600}{t_r} \quad (5)$$

t_r is the average interval time for two successive cars to continuously pass the stop line, s.

(3) Capacity of special left-turning lane

$$C_l = f_{snow-ice2} \frac{3600}{t_l} \quad (6)$$

t_l is the average interval time for two successive cars to continuously pass the stop line, s.

(4) Capacity of straight-going and left-turning mixed driving lane

$$C_{Sl} = C_l \left(1 - \frac{3}{4} \beta_l \right) f_{snow-ice1} \quad (7)$$

In the formula, β_l is the proportion of vehicles to turn left in this mixed driving lane.

(5) Capacity of straight-going and right-turning mixed driving lane

$$C_{Sr} = C_r \left(1 - \frac{1}{2} \beta_r \right) f_{snow-ice2} \quad (8)$$

In the formula, β_r is the proportion of vehicles to turn right in this mixed driving lane.

According to the definition of signalized intersection, the capacity of signalized intersection which is obtained on the basis of Method of Stopping Line is listed as follows:

$$C = \sum_{i=1}^4 \left(\sum_{j=1}^{m_i} C_{sij} + \sum_{j=1}^{m_i} C_{rj} + \sum_{j=1}^{m_{li}} C_{rlj} + \sum_{j=1}^{m_{li}} C_{slj} + \sum_{j=1}^{m_{lr}} C_{srj} \right) \quad (9)$$

In the formula, m_s , m_r , m_l , m_{sl} and m_{sr} respectively represent the number of straight-going lane, special left-turning lane, special right-turning lane, straight-left lane and straight-right one.

Based on the cross-shaped intersection and Method of Stopping Line, by analysing and systemizing mass data, and referring to the achievements of previous researches, the paper revises it according to various influencing factors in snow and ice weathers and then gives out the model to calculate the saturation flow rate of intersection lane under snow and ice conditions. To sum up, in order to determine the capacity of signalized intersection under snow and ice conditions, the author should also make clear the correction coefficient on snow and ice pavement $f_{snow-ice2}$ for the coefficient is influenced by several factors and it is also very complicated. In consequence, the paper will make detailed research about the correction coefficient of saturation flow rate under snow and ice weathers in the next section.

3 Parameter modification

3.1 DETERMINATION OF INFLUENCING FACTOR $f_{snow-ice1}$ UNDER SNOW AND ICE CONDITIONS ACCORDING TO HCM METHOD

The above part makes separate study about various factors influencing the capacity of signalized intersection under snow-ice. However, the capacity of signalized intersection under snow and ice conditions actually results from the influence of several factors, whose functions cannot be expressed by their simple product; therefore, the paper needs to consider various factors and establishes the comprehensive correction coefficient influencing the capacity of intersection under snow and

ice conditions. In details, the correction coefficient of saturation flow rate under snow and ice $f_{snow-ice1}$ is mainly influenced by weather condition Wea , lane number L_a , lane width W and lane function L_c . Thereinto, the worse the weather condition is, the smaller the correction coefficient of saturation flow rate (is); the larger the values of lane number, lane width and lane function are, the bigger the correction coefficient of saturation flow rate. As a result, the model for correction coefficient of saturation flow rate is established in Formula 10 as follows according to Gravity Model.

$$f_{snow-ice1} = k \frac{L_c^b \times W^c \times L_a^d}{Wea^a} \quad (10)$$

In the formula, a , b , c , d and k are respectively undetermined parameters.

In order to simplify the calculation process, the model is changed into linear one for treatment and the logarithm is taken towards both sides of the equal sign to obtain the following Formula 11.

$$\ln f_{snow-ice1} + a \ln Wea = \ln k + b \ln L_c + c \ln W + d \ln L_a \quad (11)$$

According to the definition, when the value of weather influencing factor Wea is 0, the value of $f_{snow-ice1}$ is 1, which means that weather factor exerts no influence to saturation flow rate. In other words, it takes the value equal to the saturation flow rate under normal conditions. Then, the paper calibrates it according to the data in field measurement and makes linear regression towards Formula 11, finally obtaining $a=1.02$, $b=0.34$, $c=0.69$, $d=-0.74$, $k=0.83$ and linear fitting $R^2=0.672$. This shows that that correction model of saturation flow rate can better explain the influence of various factors to the capacity of signalized intersection under snow and ice conditions, which is shown in Formula 12 as follows.

$$f_{snow-ice2} = \begin{cases} 0.83 \times \frac{L_c^{0.34} \times W^{0.69} \times L_a^{0.74}}{Wea^{1.02}}, & Wea \neq 0 \\ 1, & Wea = 0 \end{cases} \quad (12)$$

3.2 DETERMINATION OF INFLUENCING FACTOR $f_{snow-ice2}$ UNDER SNOW AND ICE CONDITIONS ACCORDING TO METHOD OF STOPPING LINE

As there are differences between Method of Stopping Line and HCM Method in calculation method, their coefficients under snow and ice weathers are totally different. Based on calibrating correction parameters in the model to calculate capacity according to HCM Method, the paper should also need to recalibrate corrected parameters in the capacity model based on Method of Stopping Line, for better representing the capacity of signalized intersection under snow and ice weathers.

Under snow and ice conditions, the frictions that the vehicle suffers in going straight and turning are different, which causes its acceleration of starting and braking are also different. Therefore, when calibrating these parameters, the author also needs to distinguish vehicles turning from those going straight for separate calibration. After making measurement and calculation for several times, the paper obtains calibration results of correction parameters based on Method of Stopping Line under the condition of snow and ice pavement, which is listed in Table 1 as follows.

TABLE 1 Parameter Calibration Results in Parking Line Method under Snow-Ice

Weather	Turning right	Going straight	Turning left
Normal weather	1.0	1.0	1.0
Light snow	0.74	0.78	0.73
Moderate snow	0.52	0.61	0.55
Heavy snow	0.46	0.56	0.49
Snowstorm	0.39	0.48	0.41
Road surface covered with snow	0.64	0.74	0.66
Rough road surface with ice	0.55	0.64	0.57
Smooth road surface with ice	0.50	0.54	0.48
Mixed road surface with snow and ice	0.38	0.45	0.40

Nowadays, there is rare situation where mixed driving lanes, straight-left or straight-right, exist in signalized intersection; therefore, when simulating regression, it can be ignored. According to Table 2-5, in the model based on Method of Stopping Line, snow and ice weathers exert very large influence to the capacity of road and the biggest impact can reach about 60 percent. Thereinto, the influence of straight-going lane is relatively small; left-turning lane takes the second place and the influence of right-turning lane is the biggest, which is also related with the radius of vehicles turning circle in signalized intersection. In other words, the bigger the radius is, the smaller the influence is. Furthermore, the radius of turning circle in straight-going lane approaches infinite; consequently, snow and ice weathers exert the smallest impact to it; the left-turning lane takes the second place. As the radius of turning circle in right-turning lane is the smallest, snow and ice weathers thus exert the biggest influence to it.

3.3 MODEL CALCULATION RESULT AND COMPARISON

In order to determine the precision of models previously calibrated for capacity of signalized intersection under snow and ice conditions, the paper now selects the typical “cross-shaped” intersection for investigation, then concludes and systemizes survey data for verification calculation. Then the paper makes calculation according to the data about traffic volume at the intersection between Huaihe Road and Songshan Road. Thereinto, Songshan Road-Huaihe Road is the intersection between primary and secondary trunk roads in Harbin City;

furthermore, the former possesses 8 two-way lanes, so does the latter. The place of intersection in-lane is broadened to 5 lanes. As the signal control, cycle time is 134s and the signal-timing phase is 4-phase timing, thereinto the phase of green light for going straight in south-north direction is 35s and that in east-west direction is 25s. As the left-turnings in east-west direction and south-north direction separately set up independent phases, the green light time is respectively 20s and 25s, and the yellow light time for going straight and turning left is separately 8s and 9s; right turning is not restricted by the signal light.

When the snow and ice condition is rough road surface with ice, two models are employed to respectively calculate the capacity of the intersection of Songshan Road-Huaihe Road, which are shown in Table 2. The result of model calculation is slightly smaller than that corrected from practical observation data. This is because when the signal light is yellow, the vehicles actually entering the intersection can still continue to pass the intersection; however, the model fails to calculate the capacity in this time.

TABLE 2 Result of the Intersection of Songshan Road-Huaihe Road about Capacity (HCM)

	Model calculation (pcu/h)	Corrected result from actual measurement (pcu/h)	Deviation
East in-lane	2039	2203	7.44%
South in-lane	1598	1782	10.33%
West in-lane	1994	2178	8.45%
North in-lane	1537	1639	6.22%

TABLE 3 Result of the Intersection of Songshan Road-Huaihe Road about Capacity (Method of Stopping Line)

	Model calculation (pcu/h)	Corrected result from actual measurement (pcu/h)	Deviation
East in-lane	2146	2203	2.59%
South in-lane	1673	1782	6.12%
West in-lane	2146	2178	1.47%
North in-lane	1673	1639	2.07%

According to Table 2, the difference between the result of model calculation and that of actual measurement is not large and the maximum deviation is no more than 11 percent. Therefore, the model is considered to better reflect the capacity of signalized intersection on rough road surface with snow and ice. The difference between the result of model calculation and that of actual measurement in Table 2 is not large and the maximum deviation is 12 percent. As a result, it can better reflect the capacity of each in-lane in signalized intersection on rough road surface with snow and ice.

By comparing the results from these two methods, the paper finds that the result calculated on the basis of Stopping Line Method is able to reflect the capacity of intersection, and the average deviations of intersections at Songshan Road-Huaihe Road is 3.30 percent. The difference between the result calculated on the ground of HCM Method and the capacity obtained from actual

measurement is the largest, and the average deviations in these two intersections separately reach 9.29 percent and 8.11 percent, far higher than the result obtained on the basis of Stopping Line Method. This is because the traffic flow composition of urban roads in China is complicated and the type of vehicle is various. Furthermore, the result obtained on rough road surface with ice exactly reflects the application scope of method to calculate capacity. According to the above two models, the traffic capacities of various in-lanes under various climatic conditions at intersections Songshan Road-Huaihe Road and Hongqi Street-Huaihe Road are respectively obtained from calculation and then the paper makes comparison between the result from model calculation with data in actual measurement. And the result shown in Table 4 is the result obtained on the basis of Stopping Line Method.

According to Table 4, snow and ice weathers generate obvious influence to the capacity of signalized intersection. In actual measured values, from normal weather to mixed road surface with snow and ice, the capacity of intersection has reduced by nearly 40 percent, which means that it is quite necessary to study the capacity of signalized intersection under snow and ice conditions. What's more, there is deviation between the result from model calculation established on the basis of Stopping Line Method and that from field measurement. However, the deviation keeps within the range from -7% to 12%; therefore, it is considered that the model can better represent the capacity of various in-lanes in signalized intersection under snow and ice conditions. In the end, the capacity of the whole intersection is calculated through that of each in-lane, whose result is shown in Table 5.

According to Table 5, under various weathers, the deviation between the value of model calculation and that from field observation is not large; moreover, the maximum deviation is 4.84 percent and the minimum one is 0.32 percent. This shows that the model can perfectly reflect the capacity of signalized intersection. For example, under the condition of mixed road surface with snow and ice, the traffic capacities of two intersections studied by the paper are both 6,624pcu/h in model calculation while the traffic capacities from actual observation are separately 6,477pcu/h and 6,500pcu/h. Hence, the deviation between calculated value and values in field observation are both within 200pcu/h, and the deviation ratios is separately 2.27 percent and 1.91 percent.

TABLE 4 Result of the Intersection In-lanes Capacity in Ice-snow (Method of Stopping Line)

Weather	In-lane	Songshan Road-Huaihe Road		
		Model calculation (pcu/h)	Corrected result from actual measurement (pcu/h)	Deviation
Normal weather	East in-lane	2325	2381	2.35%
	South in-lane	1824	1859	1.88%
	West in-lane	2325	2368	1.82%
	North in-lane	1824	1798	-1.45%
Road surface	East in-lane	2248	2295	2.05%
	South in-lane	1615	1822	11.36%

Zhang Yaping, Liu Chunxiao, Cheng Guozhu, Zhang Hong

with snow	West in-lane	2248	2264	0.71%
	North in-lane	1615	1738	7.08%
Rough road surface with ice	East in-lane	2146	2203	2.59%
	South in-lane	1673	1782	6.12%
Smooth road surface with ice	West in-lane	2146	2178	1.47%
	North in-lane	1673	1639	-2.07%
Mixed road surface with snow and ice	East in-lane	2058	2108	2.37%
	South in-lane	1602	1627	1.54%
with snow and ice	West in-lane	2058	1986	-3.63%
	North in-lane	1602	1576	-1.65%
with snow and ice	East in-lane	1853	1845	-0.43%
	South in-lane	1459	1439	-1.39%
with snow and ice	West in-lane	1853	1814	-2.15%
with snow and ice	North in-lane	1459	1379	-5.80%

TABLE 5 The Capacity of Signalized intersection in Ice-snow of Songshan Road-Huaihe Road and Hongqi Street-Huaihe Road (Method of Stopping Line)

Weather	Songshan Road-Huaihe Road		
	Model calculation (pcu/h)	Corrected result from actual measurement (pcu/h)	Deviation
Normal weather	8298	8406	-1.28%
Road surface with snow	7726	8119	-4.84%
Rough road surface with ice	7638	7802	-2.10%
Smooth road surface with ice	7320	7297	0.32%
Mixed road surface with snow and ice	6624	6477	2.27%

In order to show that Method of Stopping Line can better reflect the capacity of signalized intersection under snow and ice conditions, the paper now makes HCM comparison between the calculated results based on HCM Method and Conflict Point Method with the result based on Method of Stopping Line, and the results are shown in Table 6. According to Table 6, in terms of the intersection at Songshan Road-Huaihe Road, there are errors with varying degrees between the results calculated from two models and the result in actual measurement. Thereinto, the deviation between the result obtained on the basis of HCM Method and data in actual measurement is the biggest, which approaches 7 percent; the deviation between model result based on Method of Stopping Line and data in actual measurement is the smallest. However, the result is too large, which is because when the red signal light is on, the headstocks of some vehicles have surpassed the stop line, thus making the model result too large. Longitudinally speaking, the error between the results from these two methods is the biggest on road surface with snow. This means that light snow exerts the largest influence to the capacity of signalized intersection. When snow and ice weathers change to road surface with ice, the error between the result of model calculation and that in actual measurement only changes a little, which means that as snow and ice weathers change, the influence to capacity of signalized intersection mainly reflects in aspects from normal weather to road surface with snow, which is the same to the change scope of friction coefficient on road surface.

TABLE 6 The Capacity of Signalized intersection in Ice-snow

Weather	Songsshan Road-Huaihe Road				
	Model calculation (pcu/h)		Corrected result from actual measurement (pcu/h)	Deviation	
	Based on HCM Method	Based on Stopping Line Method		Based on HCM Method	Based on Stopping Line Method
Normal weather	8177	8298	8406	2.72%	-1.28%
Road surface with snow	7568	7726	8119	6.79%	-4.84%
Rough road surface with ice	7425	7638	7802	4.83%	-2.10%
Smooth road surface with ice	7107	7320	7297	2.60%	0.32%
Mixed road surface with snow and ice	6326	6624	6477	2.33%	2.27%

4 Conclusion

With snow and ice conditions as starting point, the paper calculates the traffic capacities of signalized intersection under various snow and ice conditions. Based on these two models, the author respectively calculates the traffic capacities of the whole signalized intersection at Songsshan Road-Huaihe Road and at Songsshan Road-Huaihe Road and he then compares the results with data in actual measurement. The results show that: compared with HCM Method, Method of Stopping Line possesses a smaller error with results in actual measurement, whose maximum deviation ratio is no more than 12 percent. Finally, it is considered that the capacity model based on Method of Stopping Line can well reflect the actual situation.

However, the action of taking the cross-shaped intersection for example in the study is slightly one-sided; therefore, in the future researches, the author needs to introduce the intersections in other types to make further study about the capacity of intersection under snow and ice conditions.



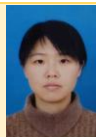

Acknowledgements

This work was financially supported by Jilin Province Science and Technology Development Project (20120747).

References

- [1] Transportation research board 2000 *Highway Capacity Manual, Special Report 209* Washington.D.C. National Research Council 778-813
- [2] Zhang Lin, Prevedouros P D 2004 Signalized intersection LOS that accounts for user Perceptions *The 83th Annual Meeting of the Transportation Research Board, Washington D.C.*
- [3] Li Dan, Shao Chunfu 2009 Signalized Intersection Level-of-Service Model Based on Fuzzy Neural Networks *Journal of Transportation Systems Engineering and Information Technology* 9(4) 127-128
- [4] LI J, Yue Z Q, Wong S C 2004 Performance evaluation of signalized urban intersections under mixed traffic conditions by gray system theory *Journal of Transportation Engineering* 130(1)
- [5] Zhang Lin 2004 *Signalized intersection level-of-service that accounts for user Perceptions* Ph.D. Dissertation, University of Hawaii
- [6] Zhang Lin 2007 Fuzzy logic-based user Perceptions of signalized intersection level of service *The 86th Annual Meeting of the Transportation Research Board, Washington D.C.*
- [7] Flannery A, Pedersen N 2005 Incorporating Customer Perceptions and satisfaction into determination of level of service *Proceeding of 84th annual Meeting of the Transportation Research Board, Transportation Research Board, Washington D.C.*
- [8] Cheng Guozhu, Mo Xuanyan, Mao Chengyuan 2011 Urban Road Traffic Safety Evaluation Method under the Condition of Ice and Snow Pavement *Journal of Transportation Systems Engineering and Information Technology* 11(1) 130-134
- [9] Liwei Hu, Yongjie Zhang 2013 Analysis on Capacity of Non-signalized Intersection on Condition of Ice-Snow *Modern Transportation, Beijing* 2013(1) 10-14
- [10] Chen Xiaoming, Shao Chunfu 2007 Influence of Bicycle Traffic on Capacity of Typical Signalized Intersection *Tsinghua Science and Technology* 2007(2) 34-38

Authors

	<p>Yaping Zhang, born in September 5, 1966, Pingxiang</p> <p>Current position, grades: Professor, Director of Transportation Engineering Department, Harbin Institute of Technology. University studies: Ph.D. in Highway and Railway Engineer (2005, Harbin Institute of Technology). Publications: 101 scientific papers. Research activities: traffic planning, highway capacity, and transport logistics.</p>
	<p>Chunxiao Liu, born in January 19, 1973</p> <p>Current position, grades: Doctoral student of transportation and management in Harbin Institute of Technology. University studies: postgraduate qualifications, master's degree, graduated from transportation planning and management of Harbin Institute of Technology. Publications: fifteen scientific research papers and ten administrative papers. Research activities: mainly engaged in the study of traffic planning and traffic management, also engaged in the study of administrative work.</p>
	<p>Hong Zhang, born in January 7, 1981</p> <p>Current position, grades: School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, student studying for doctoral degree Publications: 10 scientific papers and 3 books Research activities: transportation safety</p>
	<p>Guozhu Cheng, born in August 16, 1977, China</p> <p>Current position, grades: School of Transportation Science and Engineering, Harbin Institute of Technology, associate professor University studies: Ph.D in Road and Railway Engineering (2007, Harbin Institute of Technology) Publications: 70 scientific papers and 6 books Research activities: transportation safety</p>

Structure performance analysis of vehicular *ad hoc* networks based on complex network theory

Zhang Hong¹, Li Jie^{2*}, Lv Yuejing³

¹ School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

² School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

³ School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, Hubei 430081, China

Received 1 February 2014, www.tsi.lv

Abstract

A very promising direction in intelligent transportation system is the applications based on vehicular ad-hoc networks (VANET). VANET of a country is one of the most important indicators of its economic growth. In this paper, we analysed the characteristics of VANET network topology using complex network theory. The author contribute VANET model, analyse the statistical properties of complex network based on the degree distribution, the clustering coefficient, the average path length and find that the network has scale-free and small world features. The structure and properties of VANET have great implications for traffic congestion and urban planning. The robustness analysis indicates that VANET is robust to random attack when considering static robustness, but somewhat vulnerable to intentional attack.

Keywords: complex network, VANET, topological analysis, robustness.

1 Introduction

If we put the elements in a system as a node, and relation between the nodes as a connection, then the whole system form a network, such as computer network form interacting systems that can be viewed as a computer through cable, twisted-pair cable. There are a large number of complex systems in real-world which can be describe as network. These topology nodes abstracted from real network, known as complex network. At the end of the 1960s Paul Erdos and Alfred Renyi proposed a completely random network model (ER). However, ER random networks are theoretical constructs and cannot describe the properties of the most real-world systems. Watts and Strogatz as well as Barabási and Albert et al. proposed small world and scale-free networks [1-3]. With future development of complex network theory, several systems have been investigated using various concepts of statistical physics for complex networks, such as the average path length, the clustering coefficient, the degree distribution and betweenness. According to the existing complex network research, the high degree node is not high betweenness and vice versa. At home Li et al [4] study complex network community structure division algorithm, Gan et al [5] investigate the global adaptive synchronization problem for a class of supply chain complex networks that have nonlinearly coupled identical nodes and an asymmetrical coupling matrix. Wang et al [6] prove that a single controller can pin a coupled complex network to a homogenous solution, which is

investigated for both continuous-time and discrete-time cases. Sufficient conditions are presented to guarantee the convergence of the pinning process locally and globally. The efficiency of the derived-results is illustrated by numerical simulation. Hu et al [7] constructs China's port shipping network structure chart, and analyses the degree distribution, average path length, network density and network clustering coefficient of the network. The results show that China's port shipping network has scale-free characteristic with short average path length, high network density, and its degree distribution follows a power law. Through the computer simulation, the higher correlation coefficient also shows that data fitting is satisfactory, which further validates that China's port shipping network is a scale-free network. Complex networks have attracted increasing attention from various fields of science and engineering today. Many scholars thought the most influential nodes in communication network is more connected nodes (hubs) which is often high betweenness of nodes, but other scholars disagree with this statement [8]. Other scholars (such as Husdal, Bell, Jenelius) contacted the consequences of an event and the probability as well as taking into account the probability and consequences of a propagation event. As shown in Figure 1, the III quadrant describes the greatest probability and the most serious consequences while the I quadrant is on the contrary.

* Corresponding author - Tel: +86-027-871-95166; fax: +86-027-871-95310; E-mail: whpu302@sina.com.cn

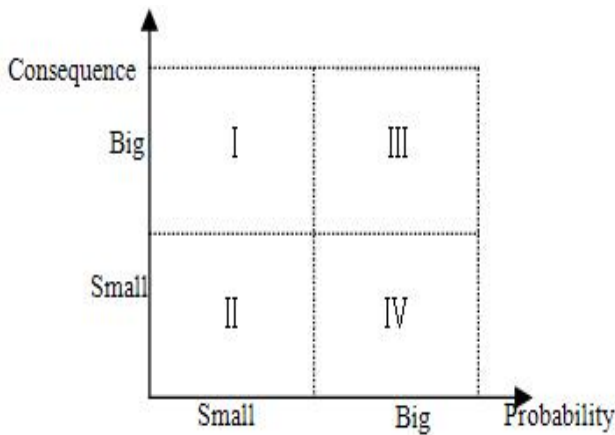


FIGURE 1 Probability-consequences matrix diagram

The vehicular ad-hoc networks (VANET) consist of a mass of nodes and connections between nodes are seriously complex, so it has the common characteristic of complex network. However, many different properties are found in VANET, such as autonomous and selective behaviours. A very promising direction in intelligent transportation system is the applications based on VANET. Through information, sharing between vehicles via VANET, the efficiency and safety of transportation system can be improved and the existing road network can give full play of its role. The high mobility of vehicles, the complex distribution of road networks and the time property of traffic peak hours lead to the highly uneven and dynamic topology of VANET. VANET is shown in the Fig. 2.

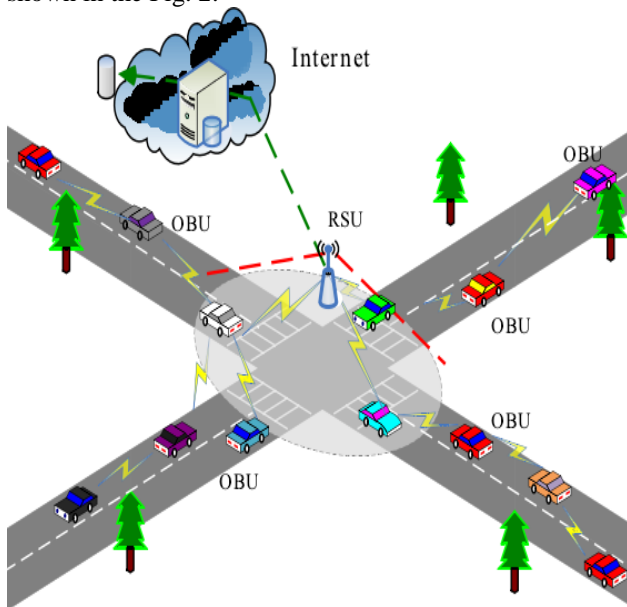


FIGURE 2 VANET network

The key problem of complex networks is to analyse their structure complexity including the topology complexity and the characteristics in different topologies, which is also a foundational theory of VANET. Moreover, it has become a commonsense that we should study the VANET systematically. In addition, adaptability and dynamic characteristics are also importation properties,

which mean that their topology structures are not fixed and unchangeable, and they will evolve with time under the external and internal force drive. Firstly, the paper introduces the fundamental principle of complex network. Some fundamental complex network models and topologies are described. Then, the research background on VANET complex network is presented and finally the special characteristic different with other complex network is described in detail.

2 Topological Properties

The basic element of complex network is nodes and edges. The basic models include regular network model topology. Under the influence of the randomness, most networks between fully regular network and fully random network show small-world property significantly. Due to less nodes included, the original network generally does not show significant scale-free property. Therefore, we take small-world network as the network morphology of the new network in this paper. In the small-world model, p denotes the random rewiring probability, which can be adopted to adjust network randomness as shown in Fig. 3.

When $p=0$, the new module shows regular network form; when $p=1$, the edges of the new network are all rewired randomly which shows ER random network form; When $0 < p < 1$, the new network is a small-world network, which lies between regular network and random network. With the increase of random rewiring probability p , the randomness of the new module increases gradually.

In VANET, nodes represent the vehicles on the platform. The edges represent connection between vehicles, which show relationship of nodes that vehicles establish contact through wireless mobile network. Same as transportation complex network, network connectivity is also the most basic requirement for VANET communication. Nodes and edges form undirected-weighted networks. The edges between nodes are bidirectional, which ignore the length between nodes in wireless networks. VANET can be abstracted as network $G(N, E)$, where N is the vehicles' number and E is the edges between the vehicles in the wireless network. If nodes represent many VANET vehicles, we use the above method to build network basic elements of nodes and the relationship between vehicles, so we get a network of VANET. The relationship between them is shown in Fig. 4.

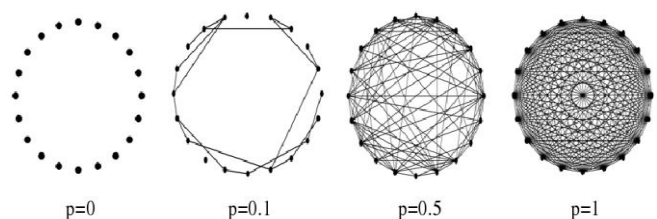


FIGURE 3 Dynamic evolution map of regular network

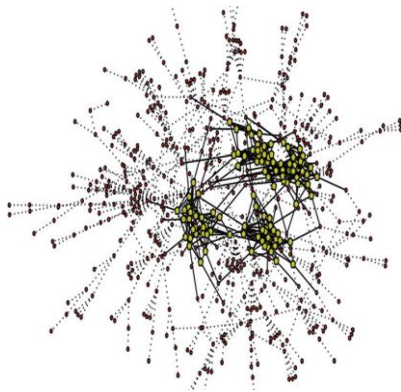


FIGURE 4 VANET network for vehicles and the relationship

Nodes represent vehicles in the VANET and the location of the nodes is not fixed. Relations between the vehicles represent whether they are wireless links. Such undirected network can use the adjacency matrix $A=(w_{ij})$. If the i and j is connected $w_{ij}=1$, otherwise $w_{ij}=0$, while the relationship of each nodes itself is 0. We get the resulting VANET network for undirected-weighted network, which the adjacency matrix is symmetric matrix. Therefore, VANET network can be expressed as a diagonal of 0 symmetrical two-dimensional matrixes. Traditionally, for network $G=[V,E]$, the adjacency matrix of the network is represented as A :

$$A = \begin{bmatrix} 0 & w_{12} & w_{13} & w_{14} \\ w_{21} & 0 & w_{23} & w_{24} \\ w_{31} & w_{32} & 0 & w_{34} \\ w_{41} & w_{42} & w_{43} & 0 \end{bmatrix}, \tag{1}$$

where w_{ij} is the weight of the network, define the edge weight as the number of vehicles' links, and the weight of node is the sum of the weights of all the edges linking with this node. In other words, the edge weight reflects the transfer convenience among network nodes. As the weighted network defined, we can investigate some vehicles, and the higher weight of node is, the more significant the vehicle is.

3 Topological Analysis of VANET

There are basic statistical parameters and degree distribution of the complex network, clustering coefficient, average shortest path length etc. [9, 10]. VANET, having complex characteristics, is a typical complex network. Based on the analysis above, we analyse the statistical properties of the network respectively.

Degree distribution: the node degrees k is the number of nodes it is directly connected to [11]. Degree of a node i is defined as $k_i = \sum_{j=1}^N a_{ij}$. All of the nodes average degree was found to be $\langle k \rangle$, which presents the network's average degrees, $\langle k \rangle = \frac{2E}{N-1} \approx \frac{2E}{N}$. In the network node, degree distribution is referred to the

probability distribution function $p(k)$. Its meaning is the probability of random nodes having k edges directly connected to, also equal to the degree k number of nodes in the total number of network nodes. If the node degree distribution follows power-law distribution, it means $p(k) \propto k^{-\gamma}$. When N is large, nodes' degree distribution is approximately for the poisson distribution, $p(k) \approx \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}$. Power-law distribution graphics

have not peak. Most of the node is only a small amount of connection, and a small number of nodes with a large number of connections. There is no characteristic scale in random networks, scholars referred to as a scale-free network. Fig. 5 and Fig. 6 show the degree distribution of regular network and Scale-Free network. They conform to Poisson distribution and power-law distribution.

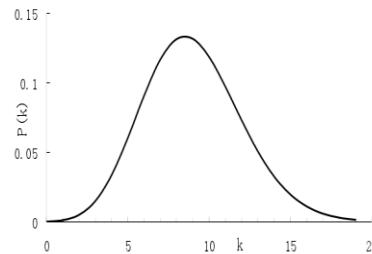


FIGURE 5 Degree distribution of regular network

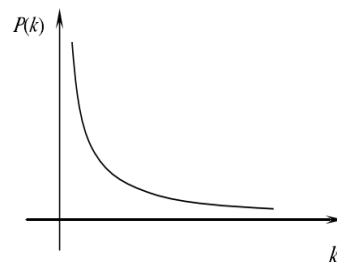


FIGURE 6 Degree distribution of Scale-Free network

Clustering coefficient: in VANET network, clustering coefficient C_i of a node is defined as the ratio of number of links shared by its neighbouring nodes to the maximum of possible links among them. For example, high clustering coefficient of the star VANET or public number VANET indicates that public pays more attention to them. The average clustering coefficient $\langle C \rangle$ reflects the VANET density relation among all nodes in the network.

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, 0 < C_i \leq 1, \tag{2}$$

$$\langle C \rangle = \frac{\sum_{i \in G} C_i}{N}. \tag{3}$$

Average path length: the distance between two nodes i and j in the network d_{ij} defined as the edges number of the shortest path of connecting two nodes. The diameter of the network was the maximum distance between two

nodes, $D=\max d_{ij}$. Network average path length L defined as the average distance between of all nodes:

$$L = \frac{2}{N(N-1)} \sum_{i,j \in G, i \neq j} d_{ij} \quad (4)$$

Studies on complex networks have found most network is a small-world network that is often characterized by high connectivity and clustering, but so far there exist few effective approaches to evaluate small-world properties. There are comparisons of random, regular and small-world network shown in Table 1.

TABLE 1 Comparison of random, regular and small world networks

Type of network	Connectivity	Clustering
Random network	Short global separation	No clustering
Regular network	Long global separation	Highly clustered
Small-world network	Short global separation	Highly clustered

4 VANET Simulation Analyses

VANET Model Based on Complex Network. Because VANET is new, specific data acquisition is difficult. In this paper, the data obtained by JAVA programming. Starting with a small number m_0 of the original node and a small number e_0 of edges. In the lower nodes, we selected 20 nodes randomly, at every time steps, add a new module of small world network containing s nodes and n edges, which become connected to m existing nodes in the network. After t time step, the total number of nodes is m_0+st and the total number of edges is $N=e_0+(m+n)t$. [12-14] Then we structure adjacency symmetric matrix w_{ij} . Using $e = m/N$ present the ratio connection number m to the vehicles' number N . Our study found that e values between 1 and 2 even if N is different. In the regular network, such as the grid structure or tree structure in the network, $e=1$. In the random network $e = (N - 1) / 2$. Studies have shown that $e > 15/7$ in random ER network. Fig. 7 represents range of the e values under different sample. In VANET, e value is between 1 and 2 which shows that VANET is different from the regular networks and random networks. It has its own features.

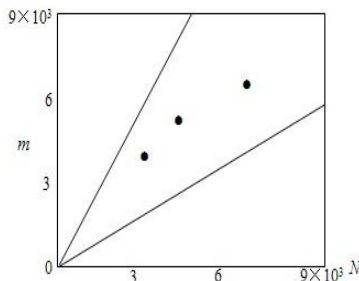


FIGURE 7 e value range under different sample scale

VANET Analysis Based on Complex Network. Degree distribution: In VANET, x present node degree, and $f(x)$ is the number of nodes. Introduce function model $f(x)=a+bx$, where a and b for the regression coefficients, and in the model a and b can be obtained by least square

method and generated through the experiment data. Remove the unreasonable data by sifting, model results to $f(x) = 9.7-0.4x$. According to the power-law distribution function $p(k) \propto k^{-\gamma}$, we take logarithm, and get degree distribution equation for deformation. A power-law distribution function, its image is intuitive. When expressed in double logarithmic coordinate axis is linear. Seen from the equation, VANET has power-law distribution characteristics. The power-law index represents the level of activity through VANET, which means vehicle joined into VANET and existing vehicle can be treated as random connections. As shown in Fig. 8, it is a certain vehicle accumulation degree distribution of the VANET.

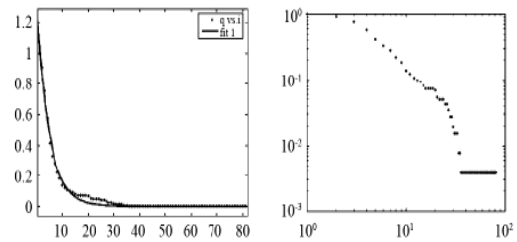


FIGURE 8 The cumulative k distribution

Clustering coefficient: VANET average aggregation coefficient $\langle C \rangle = 0.245$ is obtained by Matlab programming calculation. Compared with the poison distribution of random networks of the same scale, setting its density 0.05, the same as the density of VANET, we analyse it and get the random network clustering coefficient $C_i = 0.05$ far less than the VANET $C_i = 0.245$. Thus, VANET has larger clustering coefficient.

Average shortest path length: next we analyse the relationship matrix and get node path distribution, then the VANET average path length $L = 2.45$ be calculated by the above formula. For the huge network of VANET, it's smaller average shortest path. Average VANET vehicles can be connected to any other vehicles by three vehicles. Compared with the same scale of random network, average path length $L = 2.16$, VANET average path length is little different, as the network scale increases, the average shortest path length increase. Although the VANET number of vehicles is small, it still shows a shorter average shortest path.

The judgment of small-world characteristic: according to Watts and Strogatz analysis, to determine whether a network is small-world characteristics, It can be used an average path length L and average clustering coefficient $\langle C \rangle$ comparing with ER random network with the same scale $L_{rand}, \langle C_{rand} \rangle$. When meet the following formula (5), (6), it is the small world network.

$$L \geq L_{rand} = \frac{\ln N}{\ln \langle k \rangle} \quad (5)$$

$$\langle C \rangle \gg \langle C_{rand} \rangle = \frac{\langle k \rangle}{N} \quad (6)$$

The judgment of scale-free networks characteristic: we can judge the scale-free characteristic whether the VANET network degree distribution is power-law form. When the degree distribution is the power-law distribution, cumulative distribution function follows as:

$$P_k \propto \sum_{k'=k}^{\infty} k'^{-\gamma} \propto k^{-(\gamma-1)} \quad (7)$$

The cumulative distribution function $P_k = \sum_{k'=k}^{\infty} P(k')$ presents the probability distribution of the node's degree greater than k or equal. Therefore, we can also judge VANET scale-free network topology characteristics by plotting the cumulative distribution curve.

5 Robustness Analysis of VANET

The robustness refers to the persistence of a system's characteristic behaviour under perturbations or conditions of uncertainty. VANET robustness is referred to the ability of the network to maintain its function when the nodes or edges in the network suffer from random or intentional attack. For random attack, each node in the network is endowed with a random probability, nodes are deleted stochastically, and the ratio of the giant connected component and average path length are calculated till all nodes in the network are not connected [17]. For selective attack, the network nodes are attacked selectively, starting from the nodes with the largest connected degree, and going on according to the descending order of nodes degree. The stability performance can be represented by the relative size of S :

$$S = \frac{N'}{N} \quad (8)$$

where N and N' are respectively, the numbers of nodes of the network before and after the attack. The value of S is between $[0, 1]$. When $S=1$, the network is fully connected; when $0 < S < 1$, the network is still relatively integrated; when $S=0$, the network collapses. Starting with a small number $m_0=20$ and a small number $e_0=50$. Then according to preferential mechanism of the new nodes attaching to the entire network, we choose the nodes preferentially which would connect to the nodes existing in the VANET. We assume the probability

$$\prod(k_i) = \frac{k_i}{\sum k_i} \quad [15, 16]$$

that node I was chosen in the

module depends on its degree. Finally we obtain a VANET and $N=500$. Step 1: Compute the degree k of all nodes in the VANET; calculate the shortest path between all node pairs with Dijkstra algorithm, the distance matrix D is obtained. Step 2: Remove the node randomly and intentional according to the descending order of node degree; if there is only one of the greatest degree node, delete it directly; if there is more than one, then remove one node of them together. Step 3: Delete the q^{th} line and

q^{th} row, compute the shortest distance of node q to all other nodes i, j ; if $D(i,j) = D(i,q) + D(q,j)$, save the node pairs in array. Read node i in turn from the array; calculate the new shortest distances after removing node q ; calculate the degree k again and S . Step 4: return to step 2 and continue to carry on until the removing node rate $f=1$, then output the result.

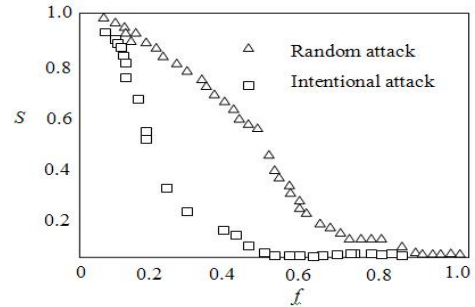


FIGURE 9 The S variation chart of VANET against random attack and intentional attack

Shown in the Fig. 9, triangle represents random attack and square is intentional attack. As seen from the analysis results, VANET shows that S drops quickly under random attack circumstance. S declines gradually, and the curve $f-S$ has two break points when f arrives 0.5431 and 0.8160 on random attack, respectively. The entire network nearly collapses when f is at 0.8792. S drops rapidly on intentional attack when $f < 0.15$, then changes a little, drops quickly when $0.18 < f < 0.35$, and falls to 0. Compared with random attack, the dropping speed of S is faster and does not appear significant fluctuation.

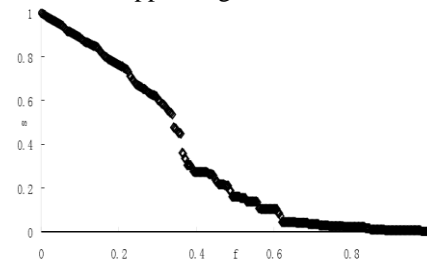


FIGURE 10 The S evolvement of VANET against random attack

Fig. 10 shows the dynamic evolvement when nodes appear random failure. It can be seen that when the network has a third of the node failure, VANET still stay connecting more than half nodes of the original network, so we can think that VANET facing random attack is robust.

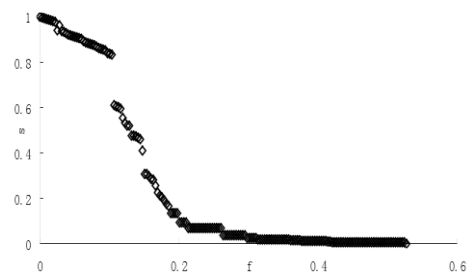


FIGURE 11 The S evolvement of VANET against intentional attack

Fig. 11 illustrates the evolution trend of the biggest connected subgraph S that attacks on node degree of

value in descending order in China VANET. The relative size of maximal connected sub-graphs S are calculated for different removing node rates f . After removing a node the degree of adjacent nodes will also be changed, therefore, in this paper, in the process of removing nodes, calculate the new shortest distances after removing node to make sure removing the maximum degree of node from the global perspective.

We can see from the Fig 11, there was an obvious two largest jumping down, the first occurred removing the 34 node, the removed node percentage is only 10.66%, and S change the size of the original 61.13%; The second occurred removing the 48 node, at this time removing 15.05% of the nodes, the relative size of network's largest connected subgraph is 30.72%; When removing the 57 node, the size of the largest network connected subgraph descended below 20%. It is clear that VANET showed a certain vulnerability to attack according to the degree descending order.

5 Conclusions

Complex network exists in many real networks. In this paper, we propose a VANET model based on complex network theory. We research the VANET topology combined with complex network. VANET model is described by the complex network method. VANET topology characteristics were analysed. Application examples show that: VANET lies between regular network and random network, and has scale-free and small-world characteristics. To find VANET network formation, understanding its development process and the way the information transmission path will provide valuable reference.

The next step we will study VANET more characteristics of complex network, and compare with other wireless complex network. There are differences between real world network and complex network. The current complex network characteristics describe the actual network, even if has two characteristics that cannot well describe the reality.

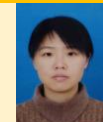
Acknowledgments

The authors wish to thank Jie Li, Zhiyun Zou, Jianzhi Gao. This work was supported in part by a grant from the National Natural Science Foundation of China (No. 51078165) and Education Department of Hubei province science and technology research program for outstanding young talent project (No. Q20131112).

References

- [1] Albert R, Barabási A L 2002 Statistical mechanics of complex networks *Reviews of modern physics* **74**(1) 47
- [2] Albert R, Jeong H, Barabási A L 2000 Error and attack tolerance of complex networks *Nature* **406**(6794) 378-382
- [3] Barabási A L, Albert R 1999 Emergence of scaling in random networks *Science* **286**(5439) 509-512
- [4] Li S, Wang X 2013 Complex Networks Community Structure Division Algorithm Based on Multi-gene Families Encoding *Journal of Computers* **8**(12) 3021-3026
- [5] Gan X, Wang J 2013 The Synchronization Problem on a Class of Supply Chain Complex Network *Journal of Computers* **8**(2) 267-271
- [6] Wang Q, Ge S, Jia L 2012 Pinning Control of Complex Network by a Single Controller *Journal of Software* **7**(10) 2258-2262
- [7] Hu B, Zong G 2013 Topology Analysis of China's Port Shipping Network *Journal of Software* **8**(10) 2581-2586
- [8] Newman M E 2003 The structure and function of complex networks *SIAM review* **45**(2) 167-256
- [9] Palla G, Derényi I, Farkas I, Vicsek T 2005 Uncovering the overlapping community structure of complex networks in nature and society *Nature* **435**(7043) 814-818
- [10] Costa L D F, Rodrigues F A, Traverso G, Villas Boas P R 2007 Characterization of complex networks: A survey of measurements *Advances in Physics* **56**(1) 167-242
- [11] Zakharov P 2007 Diffusion approach for community discovering within the complex networks: LiveJournal study *Physica A: Statistical Mechanics and its Applications* **378**(2) 550-560
- [12] Soh H, Lim S, Zhang T, Fu X, Lee G K K, Hung T G G, Wong L 2010 Weighted complex network analysis of travel routes on the Singapore public transportation system *Physica A: Statistical Mechanics and its Applications* **389**(24) 5852-5863
- [13] Zou Z, Xiao Y, Gao J 2013 Robustness analysis of urban transit network based on complex networks theory *Kybernetes* **42**(3) 383-399
- [14] Paul G, Tanizawa T, Havlin S, Stanley H E 2004 Optimization of robustness of complex networks *The European Physical Journal B-Condensed Matter and Complex Systems* **38**(2) 187-191
- [15] Zou Z, Lai J, Gao J 2013 Reducing the vulnerability of network by inserting modular topologies *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**(1)

Authors



Zhang Hong, born in January 1981, Inner Mongolia

Current position, grades: studying for doctoral degree within the School of Civil Engineering and Mechanics in the Huazhong University of Science and Technology, whose tutor is Professor Li. She has participated in several transportation programs and has done research under the guidance of Professor Li.



Li Jie, born in November 1954, Xianning, Hubei Province

Current position, grades: Professor within the School of Civil Engineering and Mechanics in the Huazhong University of Science and Technology.
Scientific interest: transportation planning and management, and traffic network analysis. He is presided over the project supported by National Soft Science Research Project (210GXS5D237) and many projects of Hubei province.
Publications: His scientific activity has materialized in over 50 articles and scientific papers.



Lv Yuejing, born in December 1979, Harbin, Heilongjiang Province

Current position, grades: is a student and teacher within the School of Automobile and Traffic Engineering in Wuhan University of Science and Technology.
Scientific interest: engaged in transportation planning and management, and road traffic.
Publications: 5 research papers in several journals.

A fast on-line two-dimensional sizes measurement method for micro part

Xingping Wei^{*}, Jun Li, Xiaoyu Liu, Shiping Zhao

School of Manufacturing Science and Engineering, Sichuan University, Chengdu, SiChuan, China, 610065

Received 8 October 2013, www.tsi.lv

Abstract

In order to meet the manufacturing process of micro part, where the space is small, the production speed, temperature and vibration frequency are high, a novel method, which is used to measure the two-dimensional (2-D) sizes of the micro part fast and on-line, was proposed based on optical fibre image bundles and CCD camera. Double parallel lights were projected on the part symmetrically. The outline information of the part image on CCD camera through optical fibre image bundles and the part 2-D sizes can be obtained from the CCD camera. The optical fibre image bundles can be embedded into tool-set or fixed in limited space due to its small size, flexibility and bendability. Therefore, the method overcomes the bad influences from the part hard processing conditions effectively. Experiments were carried out to measure workpiece with dimensions of 6.124 mm x 0.424 mm. The results show that the measurement time was less than 0.2 second and the accuracy was up to 25µm.

Keywords: On-line Measurement, Micro-part, Optical Fibre Image Bundles, Two-dimensional Sizes, CCD

1 Introduction

Micro parts have been used widely in recent years due to its various advantages such as less consumption of energy and material, low costs, small size and light-weight and so on. The typical fields of the application of the micro parts include the industries of military, automobile, electric, aerospace as well as medical equipment. There are all kinds of techniques which can be used for the manufacture of micro parts. Micro-manufacturing is mainly categorized as micro-electromechanical systems (MEMS)-based manufacturing and non-micro-electromechanical systems (non-MEMS)-based manufacturing. MEMS-based manufacturing involves techniques such as photolithography, LIGA, plating, chemical etching, laser fabrication, etc. While non-MEMS-based manufacturing often involves techniques such as mechanical machining, injection molding, forging, extrusion, stamping etc. [1]. Although fundamental issues relates to materials, processes and analysis of micro-manufacturing have been studied intensively in past two decades [2-7]. But still have some challenges on the dimensional measurement for micro part, especially fast on-line dimensional measurement. Firstly, compared with the traditional macro-manufacturing, the micro-manufacturing available space is smaller, so location of sensors is a very difficult task. Secondly, the production speed of the micro-manufacturing is fast. Some equipments manufacture's speed up to 1000 workpieces per minute. Thirdly, the

temperature and vibration frequency of the micro-manufacturing processing site are often high [8].

Currently, many techniques of micro part dimensional sizes measuring have been proposed. Yang etc. have developed a high precision micro coordinate measuring machines (M-CMM) with measurement uncertainty up to 50 nm in a measuring volume of 30mm x 30mm x 10mm (XYZ) [9]. Tosello G etc. use optical coordinate measuring machines (OCMM) to obtain the dimensional of micro channels and the surface roughness [10]. S Ontiveros proposes a dimensional measurement method of micro moulded parts by computed tomography (CT) [11]. In addition, scanning electron microscopy (SEM), atomic force microscopy (AFM) and scanning tunnelling microscopy (STM) are all widely used in micro and nano part surface analysis techniques capable of very high accuracies [12].

However, all these methods are off-line measurement. Each of them demands large space for allocation and enough time to analysis the data. To the best of our knowledge, it is not a fast on-line two-dimensional sizes measurement method for micro part has been reported.

In order to meet the manufacturing process of micro part, especially (non-MEMS)-based manufacturing, where the space is small, the production speed and vibration frequency are high, we propose a novel fast on-line 2-D sizes measurement method for micro part in this paper. Double parallel lights are projected on the micro part symmetrically, passing through optical fibre image bundles and the telecentricity modules at its both ends. The micro part 2-D sizes information can be obtained

* Corresponding author - Xingping Wei. Tel: +86-028-8772-0507 fax: +86-028-8772-0507; E-mail: weixingping314@163.com

from the CCD camera. The front-end probe consists of telecentric lens and the optical fibre image bundles, both of them volume is smaller significantly compared with the normal camera, and the optical fibre image bundles have the merits of flexible, bendable and long distance signal transmission. Therefore, the front-end probe can be embedded into tool-set or fixed in limited space. This paper is organized as follows. Section 2 gives a basic measurement principle of 2-D shape for micro part based on the optical fibre image bundles. In section 3 the measurement experiment and results of the micro part 2-D shapes are presented. Section 4 the resolution and mainly error sources of the system are analysed. Finally,

the conclusion and further study are also described in the end.

2 Principle of the measurement

The fast on-line measurement system of 2-D shape for micro part consists of double lights, two telecentric lens, optical fibre image bundles, charge coupled device (CCD), frame grabber and computer, as shown in FIGURE 1. The 2-D contour of micro part is imaged on the CCD through the optical fibre image bundles, then frame grabber and image acquisition software process the data and the 2-D sizes of micro part can be obtained.

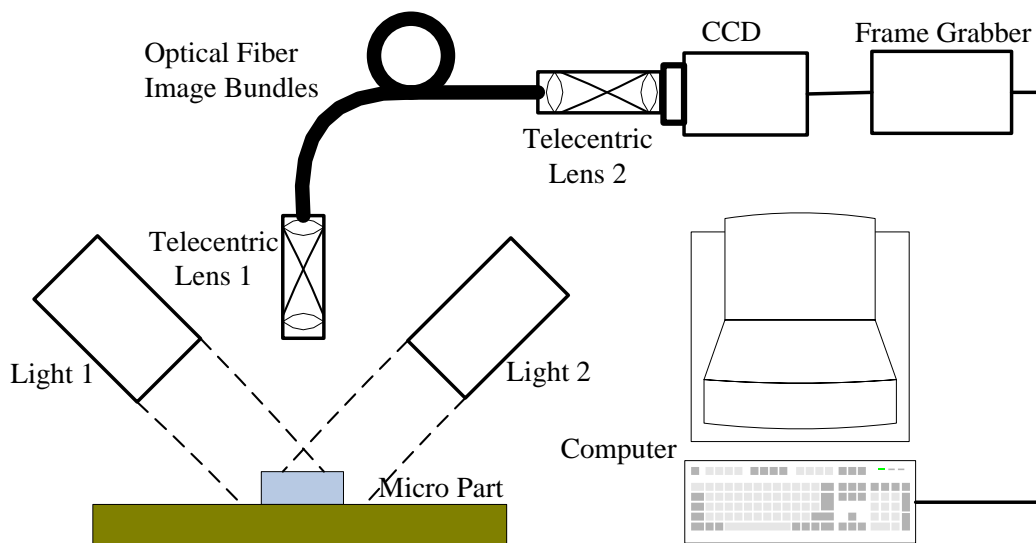


FIGURE 1 Schematic diagram of the measurement system

2.1 THE OPTICAL FIBER IMAGE BUNDLES AND TELECENTRIC LENS

Optical fibre image bundles have the merits of flexible, bendable and long distance image-transmission, so it used a wide range of applications, including medicine for endoscope, industrial remote vision systems, ordered array detectors, hazardous environment imaging, defence and research. The major technique parameters of the optical fibre image bundles used in this paper as shown in table 1.

TABLE 1 The major parameters of the optical fibre image bundles

Item	Parameters	Unit
Quality area	6 x 6	mm
Single core fibre size	17 (element 6x 6array)	um
Numerical aperture	0.56	
Resolution	34	Lp/mm
Temperature resistance	-40 to +125	°C
Diameter	12	mm
Length	800	mm

Telecentric lens have the unique property of purely orthographic projections of scene points and maintaining a constant magnification over a specific range of object distances, and show very low distortion degree, normally

less than 0.2%. So they are being widely used in many machine vision applications especially for accurate dimensional measurement of 2-D parts and components of different heights [13]. In this paper, the size and weight of the telecentric lens 1 decide the front-end probe whether it can be installed in compact space. The major parameters of the telecentric lens 1 as shown in table 2.

TABLE 2 The major parameters of telecentric lens 1

Item	Parameters	Unit
Maximum diameter	20	mm
Length	43	mm
Weight	40	gram
Magnification	0.75X	
Distortion	Less than 0.2%	
Depth of field	16	mm

We can find that the volume and weight of telecentric lens 1 is reduced significantly compared with the normal camera (In this paper, the type of camera is WAT-535EX2. The length, width and height are 53.5mm, 43.5mm and 44mm, respectively, and the weight is approximate 125 gram.) The schematic diagram of the optical fibre image bundles mounting with telecentric lens shown as in figure 2.

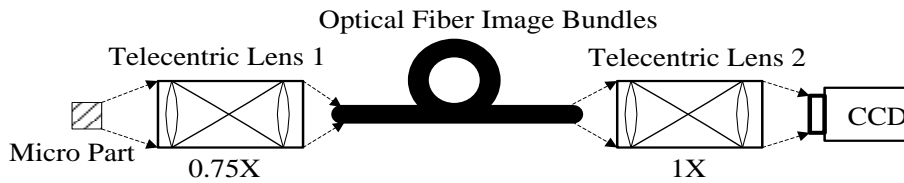


FIGURE 2 Schematic diagram of the optical fibre image bundles mounting with telecentric lens

2.2 ILLUMINATION OF THE MEASUREMENT SYSTEM

A singular light might meet some problems in practical measurement system, such as blind area and uneven illumination, etc. Symmetrically placed double lights can deals with them and the measurement precision can be improved. The telecentric lens 1 is vertically fixed upon the part and two parallel lights are symmetrically placed on both sides of it, as shown in figure 1. The illumination condition of the measurement system will be improved by placed on double lights.

3 Experiment and results

To verify this method performance, an experiment system is established to evaluate the accuracy of the measurement system. Besides the earlier mention that the optical fibre image bundles mounting with two telecentric lens to image micro part on the CCD, the WAT-535EX2 camera with resolution 768 x 494 and MV4000 frame grabber are employed. The size of a pixel is 6.35um x 7.4um. We use a pin of Microcontroller Unit (MCU, the type is STC89C51RC+) which the pins packaging with dual-in-line for measurement. The outside view of the MCU is as shown in figure 3. The pins are made by micro stamping. A pin dimension was measured through ZYM-500GS measuring microscope, the length and maximum width of the measurand are 6.124mm and 0.424mm, respectively, as shown in figure 4.



FIGURE 3 The outside view of the MCU

The micro part is placed at the working distance of the telecentric lens1, and an image is taken by the frame grabber, then software Matlab combined with VC++ to process the obtained image. The boundary of the micro part is extracted as using image processing algorithms Canny with threshold 0.6, as shown in FIGURE 5. The measurand has been measured ten times repeatedly. Time consumption is less than 0.2 second in every time,

and the experiment results are shown in TABLE 3. It is clear that both of the length and maximum width of the micro part errors are less than 25um.

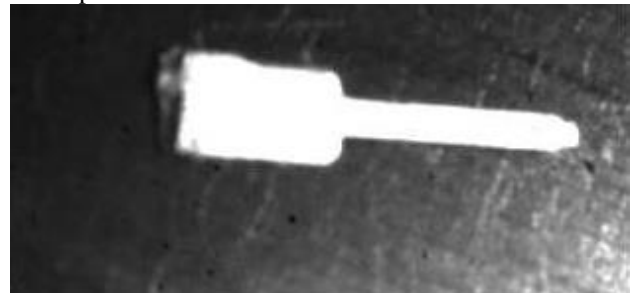


FIGURE 4 The micro part (measurand)

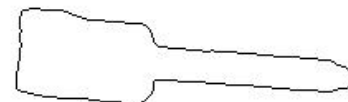


FIGURE 5 Boundary of the micro part

TABLE 3 Experiment results (unit: mm)

Number	Length(6.124)	Error	Maximum width(0.424)	Errors
1	6.103	-0.021	0.413	-0.011
2	6.148	0.024	0.408	-0.016
3	6.140	0.016	0.437	0.013
4	6.115	-0.009	0.415	-0.009
5	6.129	0.005	0.417	-0.007
6	6.118	-0.006	0.433	0.009
7	6.132	0.008	0.420	-0.004
8	6.128	0.004	0.411	-0.013
9	6.104	-0.020	0.430	0.006
10	6.126	0.002	0.439	0.015

4 Discussion

4.1 RESOLUTION OF THE MEASURE SYSTEM

The system resolution is mainly restricted by following factors, resolution of the optical fibre image bundles and CCD, magnification of the telecentric lens 1 and the telecentric lens 2. The optical fibre image bundles resolution is principally depend on the diameter of single core fibre and it arrangement mode. There usually are hexagonal and square arrangement mode, the corresponding calculate their resolution formulas are described by equations (1) and (2), respectively.

$$r_{oh} = \frac{\sqrt{3}}{2} d, \quad (1)$$

$$r_{os} = d, \quad (2)$$

where d is diameter of the single core fibre, r_{oh} and r_{os} are resolution of the hexagonal and square arrangement mode, respectively. In this paper, the core fibres are arranged by hexagonal mode, the diameter of the single core fibre is 17um, according to equation (1), we have r_{oh} is about equal to 14.7um.

The solution of the CCD is described by

$$r_c = \max\left\{\frac{W}{t_{pw}}, \frac{H}{t_{ph}}\right\}, \quad (3)$$

where W and H are the width and height of the CCD, t_{pw} and t_{ph} are the total effective pixels on the width and height direction, respectively. The WAT-535EX2 camera dimensions is 1/3 inch, the $W \times H$ of the CCD are 4800um \times 3600um, the total effective pixels on the width and height direction are 768 \times 494, so r_c is about equal to 7.3um.

As mentioned above, the image of the measurand is coupled into the optical fibre image bundles through telecentric lens 1, then imaging on the CCD through telecentric lens 2. We set the magnification of the telecentric lens 1 and telecentric lens 2 are m_1 and m_2 , respectively. Here the value of m_1 is equal to 0.75X and m_2 is equal to 1X. So the resolution of this measure system is expressed by the equation (4):

$$r = \max\left\{\frac{m_2 r_{oh}}{m_1}, r_c\right\}. \quad (4)$$

Combining equations (1) to (4), the resolution of the whole system can be acquired, and the value is equal to 19.6um. It is obvious that the resolution of the system is mainly impacted by the optical fibre image bundles and magnification of the telecentric lens 1 and telecentric lens 2.

4.2 MAIN ERROR SOURCES OF THE MEASURE SYSTEM

There are mainly three aspects contributing to the error of the measure system, including image-transmission errors of the optical fibre image bundles, distortion of the telecentric lens and CCD noises. The principal error of system is caused by the first one, low resolution of the optical fibre image bundles lead to the result that obtained image is not clear, the arrangement errors and break of the some single fibres also cause the errors. The

distortions of the telecentric lens including radial, decentring and thin prism distortions. The first one is caused by imperfect lens shape and manifests itself by radial positional error only, while the second and third types of distortions are generally caused by improper lens assembly, and the distortion of the telecentric lens is less than 0.2 percent. Photo noise, dark current noise jamming signal have influence on the resolution of the CCD image, which also reduces the measure precision of the system.

5 Conclusions

Manufacturing process of micro part features limited space, high of the production speed, temperature and vibration frequency makes fast on-line 2-D sizes measurement using common method difficult. This paper proposes a new method of fast on-line two-dimensional sizes measurement to solve this problem, and the corresponding equipment has been designed. Different from the conventional machine vision measurement system, CCD camera through optical fibre image bundles with two telecentric lens obtained image is presented. The volume and weight of the front-end probe is reduced significantly compared with only taking CCD camera directly. Experiment results prove the errors are less than 25um and time consumption is less than 0.2 second. Thus, the method is very suitable for fast on-line two-dimensional sizes measurement with space limited, especially fits well for micro-manufacturing. Although this method can solve the problem of fast on-line 2-D sizes measurement for micro part, it is not limited to this area. According to its principle, the method may also be used to measure other objects with demand for small space and high speed. The drawbacks to this method are as follows: (1) the precision of the measurement system is mainly depended on the precision of optical fibre image bundles and the magnification of telecentric lens, and improve of them will cost expensive. (2) The threshold is used in extract the boundary of micro part, at present, the threshold is confirmed through many times test that is effectively distinguishing the border.

Future work must address the above drawbacks. The higher resolution of optical fibre image bundles and bigger magnification of telecentric lens will be employed to improve the precision of the system. Further, more micro part will be tested in order to conclusion about the rule of threshold confirm. Finally, optimize the image processing algorithms to reduce the total time consumption.

Acknowledgements

The authors would like to appreciate the financial support from the Doctoral Program of Higher Education of China grant 20110181110084.

References

- [1] Qin Y, Brockett A, Ma Y, Razali A, Zhao J, Harrison C, Pan W, Dai X, Loziak D 2010 Micro-manufacturing: research, technology outcomes and development issues *Int J Adv Manuf Technol* **47**(1) 821-37
- [2] Qin Y 2006 Micro-forming and miniature manufacturing systems—development needs and perspectives *Journal of Materials Processing Technology* **177**(1) 8-18
- [3] Vollerston F, Hu Z, Schulze Niehoff H, Theiler C 2004 State of the art in microforming and investigations into micro-deep drawing *Journal of Materials Processing Technology* **151**(5)70-9
- [4] Geiger M, Kleiner M, Eckstein R, Tiesler N, Engel U 2001 Microforming *Ann. CIRP* **50** 445–62
- [5] Yang M S, Li Y and Yuan Q L 2013 A hybrid method to deformation force of high-speed cold roll-beating forming *Journal of Digital Information Management* **11** 146-53
- [6] Li W Y, Shi S X, Wang F F, Zhang Z H, Ma T J, Li J L 2012 Numerical simulation of friction welding processes based on ABAQUS environment *Journal of Engineering Science and Technology Review* **5**(3) 10-19
- [7] Qin Y 2010 *Micro-Manufacturing Engineering and Technology* Oxford Univ. Press: Oxford Chapter 22
- [8] Yang P, Tomohiko Takamura, Satoru Takahashi, Kiyoshi Takamasu, Osamu Sato, Sonko Osawa, Toshiyuki Takatsuji 2011 Development of high-precision micro-coordinate measuring machine: Multi-probe measurement system for measuring yaw and straightness motion error of XY linear stage *Precision Engineering* **35** 424-30
- [9] Tosello G, Marinello F and Hansen H N 2012 Characterisation and analysis of microchannels and submicrometre surface roughness of injection moulded microfluidic systems using optical metrology *Plastics Rubber and Composites* **41**(5) 29-39
- [10] Ontiveros S, Yag`ue-Fabra J A, Jim`enez R, Tosello G, Gasparin S, Pierobon A, Carmignato S, Hansen H N 2012 Dimensional measurement of micro-moulded parts by computed tomography *Measurement Science and Technology* **23** 1-9
- [11] D Sarid 1991 *Scanning Force Microscopy* Oxford Univ. Press: Oxford
- [12] Li D, Tian J D 2013 An accurate calibration method for a camera with telecentric lenses *Optics and Lasers in Engineering* **51** 538–41

Authors



Wei Xingping, born in February 1982, SanDou County, GuiZhou Province, P.R. China

Current position, grades: PHD Candidate of School of Manufacturing Science and Engineering, SiChuan University, China.

University studies: received his B.Sc. in Measurement and Control Technology and Instrumentation Program from SiChuan University in China. He received his M.Sc. from XiHua University in China.

Scientific interest: His research interest fields include measurement & control, machine vision, signal/image processing and optics.

Publications: 8 papers published in various journals.

Experience: He has participated six scientific research projects.



Li Jun, born in October 1983, JiangYou County, SiChuan Province, P.R. China

Current position, grades: PHD Candidate of School of Manufacturing Science and Engineering, SiChuan University, China.

University studies: He received his in Engineering M.Sc. from Southwest University of Science and Technology in China.

Scientific interest: His research interest fields include Electrical Discharge Machining, Micro Electrical Discharge Machining and Artificial Intelligence.

Publications: 6 papers published in various journals.

Experience: He has participated three scientific research projects.



Liu Xiaoyu, born in June 1987, Inner Mongolia Province, P.R. China

Current position, grades: PHD Candidate of School of Manufacturing Science and Engineering, SiChuan University, China.

University studies: She received her M.Sc. from SiChuan University in China.

Scientific interest: Her research interest fields include manufacturing and measurement technology of micro manufacturing.

Publications: 10 papers published in various journals.

Experience: She has participated five scientific research projects.



Zhao Shiping, born in December 1957, ChongQing Municipality, P.R. China

Current position, grades: the Professor of School of Manufacturing Science and Engineering, SiChuan University, China.

University studies: He received his PHD in Precision Instruments and Machinery from ChongQing University in China.

Scientific interest: His research interest fields include Measurement and Control Technology, Robot Technology and signal/image processing etc.

Publications: more than 50 papers published in various journals.

Experience: He has teaching experience of 23 years, has completed twenty scientific research projects.

Basic set of equations in mesoscale meteorological model associated with local fractional derivative operators involving the cantor and cantor-type spherical coordinates

Meng Xian-Yong^{1, 2, 3}, Liu Yang^{1, 2}, Gao Yu-Xiao¹, Wang Mei¹, Liu Zhi-Hui^{1*}

¹ School of Resources and Environment Science, Xinjiang University, Urumqi, 830046, China

² Key Laboratory of Oasis Ecology Ministry of Education, Xinjiang University, Urumqi, 830046, China

³ Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang, 830011, China

Received 1 March 2014, www.tsi.lv

Abstract

Recently, the local fractional calculus theory was applied to process the non-differentiable phenomena in fractal domain. The main object of this paper is to present the basic set of equations in mesoscale meteorological model on the Cantor sets involving local fractional derivative operators and the corresponding cantor-type spherical coordinate equations. It is shown that these equations are efficient and accurate for describing some of atmospheric motion.

Keywords: Mathematic, basic set of equations, mesoscale meteorological model

1 Introduction

Recently, the local fractional calculus theory was applied to process the non-differentiable phenomena in fractal domain (see [6–11] and the references cited therein). There are some local fractional models, such as the local fractional Fokker-Planck equation [6], the local fractional stress-strain relations [7], the local fractional heat conduction equation [11], wave equations on the Cantor sets [12], and the local fractional Laplace equation [20].

In the different process of atmospheric motion, many problems of dynamics had already been described by a number of classic basic equations, the description of these equations establish on the preferable smoothness. However, in the different process of atmospheric motion, it exists both large scale turbulence and small scale turbulence. Therefore, the problems of dynamics is no longer can be describe by classic basic equation due to the turbulent flows may be of fractal character. Meanwhile, it has relevance between the fractional order calculus and fractal. Thus, it can use the operator with fractional order of gradient, divergence etc. To generalize the atmospheric dynamic equation in the corresponding fractal form.

The main aim of this paper is present in the mathematical structure of the basic set of equations in mesoscale meteorological model in local fractional derivative and to propose their forms in the Cantor-type spherical coordination.

2 Mathematic Tools

2.1 IN THE CARTESIAN COORDINATE SYSTEM

In this part, we will introduce the local fractional derivative. It is need to bring in local fractional derivative, which define as:

$$D_{0+}^{\alpha} f(x_0) = \frac{d^{\alpha} f(x)}{dx^{\alpha}} \Big|_{x=x_0} = \lim_{x \rightarrow x_0} \frac{\Delta^{\alpha} [f(x) - f(y)]}{(x - x_0)^{\alpha}}, \quad (1)$$

where in,

$$\Delta^{\alpha} (f(x) - f(x_0)) \cong \Gamma(1 + \alpha) \Delta (f(x) - f(x_0)). \quad (2)$$

For a function of three variables, the vector form can be written in the form Ref. [2]

$$\gamma(x, y, z) = L(x, y, z)e_1^{\alpha} + M(x, y, z)e_2^{\alpha} + N(x, y, z)e_3^{\alpha}. \quad (3)$$

Let $u(x, y, z) = u_1(x, y, z)e_1^{\alpha} + u_2(x, y, z)e_2^{\alpha} + u_3(x, y, z)e_3^{\alpha}$ a local fraction vector field and $\varphi(x, y, z)$ is a local fractional scalar field, the computing rules of Hamilton operator are valid as follows Ref. [2].

(1) The local fractional gradient operator defined as [2]:

$$\nabla^{\alpha} \varphi = \frac{\partial^{\alpha} \varphi}{\partial x_1^{\alpha}} e_1^{\alpha} + \frac{\partial^{\alpha} \varphi}{\partial x_2^{\alpha}} e_2^{\alpha} + \frac{\partial^{\alpha} \varphi}{\partial x_3^{\alpha}} e_3^{\alpha}. \quad (4)$$

* Corresponding author - Tel: +86-135-792-91801; fax: +86-135-792-91801; E-mail: lzh@xju.edu.cn

The local fractional divergence and curl of local fractional vector field are written in the form [2]:

$$\operatorname{div}^\alpha u = \nabla^\alpha u = \frac{\partial^\alpha u_1}{\partial x_1^\alpha} + \frac{\partial^\alpha u_2}{\partial x_2^\alpha} + \frac{\partial^\alpha u_3}{\partial x_3^\alpha}, \quad (5)$$

$$\operatorname{curl}^\alpha u = \nabla^\alpha \times u = \left(\frac{\partial^\alpha u_3}{\partial x_2^\alpha} - \frac{\partial^\alpha u_2}{\partial x_3^\alpha} \right) e_1^\alpha + \left(\frac{\partial^\alpha u_1}{\partial x_3^\alpha} - \frac{\partial^\alpha u_3}{\partial x_1^\alpha} \right) e_2^\alpha + \left(\frac{\partial^\alpha u_2}{\partial x_1^\alpha} - \frac{\partial^\alpha u_1}{\partial x_2^\alpha} \right) e_3^\alpha. \quad (6)$$

2.2 IN THE SPHERICAL COORDINATE SYSTEM

The spherical coordinate in the three-dimensional space is the form (λ, φ, r) , in the classic differential, the gradient is:

$$\nabla = i \frac{1}{r \cos \varphi} \frac{\partial}{\partial \lambda} + j \frac{1}{r} \frac{\partial}{\partial \varphi} + k \frac{\partial}{\partial r}. \quad (7)$$

It can be generalized to the fractional calculus, the gradient of this definition is:

$$\nabla^\alpha = i^\alpha \frac{1}{r^\alpha \cos^\alpha \varphi} \frac{\partial^\alpha}{\partial \lambda^\alpha} + j^\alpha \frac{1}{r^\alpha} \frac{\partial^\alpha}{\partial \varphi^\alpha} + k^\alpha \frac{\partial^\alpha}{\partial r^\alpha}. \quad (8)$$

Similarly, it can get the divergence of this definition, like:

$$\operatorname{div}^\alpha u = \nabla^\alpha u = \frac{1}{r^\alpha \cos^\alpha \varphi} \frac{\partial^\alpha u_2}{\partial \lambda^\alpha} + \frac{1}{r^\alpha \cos^\alpha \varphi} \frac{\partial^\alpha (u_\varphi \cos^\alpha \varphi)}{\partial \varphi^\alpha} + \frac{1}{r^2} \frac{\partial^\alpha (u_r r^2)}{\partial r^\alpha}. \quad (9)$$

For above all, the following equations are valid:

$$\operatorname{div}^\alpha u = \nabla^\alpha u = \frac{\partial^\alpha u_1}{\partial x_1^\alpha} + \frac{\partial^\alpha u_2}{\partial x_2^\alpha} + \frac{\partial^\alpha u_3}{\partial x_3^\alpha}, \quad (10)$$

$$\nabla^\alpha (uv) = (\nabla^\alpha u)v + u\nabla^\alpha v. \quad (11)$$

2.3 THE INTERAL OF FRACTIONAL ORDER

(1) The local fractional line integral of the function $u(x_p, y_p, z_p)$ in the local fraction vector form.

$$u(x_p, y_p, z_p) = u_1(x_p, y_p, z_p) e_1^\alpha + u_2(x_p, y_p, z_p) e_2^\alpha + u_3(x_p, y_p, z_p) e_3^\alpha. \quad (12)$$

Along a fractal line l^α is written as [2]:

$$\int_{l^\alpha} u(x_p, y_p, z_p) dl^\alpha = \lim_{N \rightarrow \infty} \sum_{p=1}^N u(x_p, y_p, z_p) \Delta l_p^\alpha, \quad (13)$$

where for $p = 1, 2, \dots, N$ and N elements of line Δl_p^α , it is required that all $|\Delta l_p^\alpha| \rightarrow 0$ as $N \rightarrow \infty$.

(2) The local fractional surface integral of the given function across a surface $s^{(\beta)}$ is defined as [2]:

$$\iint_{s^{(\beta)}} u(x_p, y_p, z_p) ds^{(\beta)} = \lim_{N \rightarrow \infty} \sum_{p=1}^N u(x_p, y_p, z_p) n_p \Delta s_p^{(\beta)}, \quad (14)$$

where for $p = 1, 2, \dots, N$ and N elements of area $\Delta s_p^{(\beta)}$ with a unit normal local fractional vector n_p , it is required that all $\Delta s_p^{(\beta)} \rightarrow 0$ as $N \rightarrow \infty$.

(3) The local fractional volume integral of the given function in a fractal region $v^{(r)}$ is given by [2]:

$$\iiint_{v^{(r)}} u(x_p, y_p, z_p) dv^{(r)} = \lim_{N \rightarrow \infty} \sum_{p=1}^N u(x_p, y_p, z_p) \Delta v_p^{(r)}, \quad (15)$$

where for $p = 1, 2, \dots, N$ and N elements of volume $\Delta v_p^{(r)}$, it is required that all $\Delta v_p^{(r)} \rightarrow 0$ as $N \rightarrow \infty$.

Let us consider a local fractional vector field $u = u_1 e_1^\alpha + u_2 e_2^\alpha + u_3 e_3^\alpha$, the following result hold [27].

(4) Divergence Theorem of local fractional field states that:

$$\iiint_{v^{(r)}} \nabla^\alpha u dv^{(r)} = \iint_{s^{(\beta)}} u ds^{(\beta)}. \quad (16)$$

3 Peculiar Properties of Fractional Operator

3.1 THE PROPERTY OF MEMORY

The integral of fractional order has a character of memorability. The role of memory functions for compliance retardation and modules relaxation in viscoelastic materials is examined. The complexity of viscoelastic materials that occurs in the linear domain was explained by the influence of modelling these effects using the fractional calculus, such as Heat equation with memory is established, under some general and reasonable conditions in Ref. [17].

At this point, we can use a brief description to explain the memorability of fractional order, take Newtonian equation as an example, for the particle of unit mass,

$$m \frac{dV}{dt} = F(r, v, t). \quad (17)$$

In this expression, r as displacement vector, v as speed, t as time, transform the expression (14), it can get a new expression like below:

$$m \frac{dv}{dt} = \int_0^r k(t-\tau) F(r, v, t) dt \quad (18)$$

In the expression, the $k(t-\tau)$ is given kernel function with memory. The characteristic of non-locality in (17) can be obtained from locality. Where in the retardation effects are taken into account, this kind of equations can open new possibilities of understanding the classical mechanics.

3.2 THE PROPERTY OF GENERALIZED FLUX

The traditional diffusion motions use the second order convection - The classical case of Fick's second law equation - to describe, like below:

$$\frac{\partial C(x, t)}{\partial t} = D \frac{\partial^2 C(x, t)}{\partial x^2} \quad (19)$$

In the expression, $C(x, t)$ is the particle concentration in the space x at the time t ; v is convection velocity; D is diffusion coefficient.

The characteristic of Fick's law is local diffusion, Flux J which in the space of a point is proportional to the concentration gradient in a small area, however, The Fick's law don't consider the affection from particle migration by other particle, also, without considering the impact from history.

However, in the complicated system, the particle movement in different time and the particle movement in different spatial point have influence each other, therefore, when it study the particle movement at one time in one spatial point, it is need to consider the influence of particle movement at other time in other spatial point.

Thus, it is need to consider the relevancy from spatially and temporally respectively, but do not consider the spatially and temporally coupling effect of spatially and temporally, then use the limitation processing method to process the traditional second-order diffusion equation in spatially and temporally, and get the fractional order fluxional anomalous diffusion equation.

In case of one-dimensional, the flux expression of limited particle can be modified to the equality of the relation between particle flux and particle concentration [4].

$$\int_0^r d\tau \int_0^x J(x', t) dx' = \int_0^r d\tau \int_0^x k(x, x'; t, \tau) C(x', \tau) dx' \quad (20)$$

In this expression, $k(x, x'; t, \tau)$ is the kernel function with the property of diffusion, due to it is not need to consider the coupling interaction of spatial and temporal,

the kernel function with the property of diffusion can be express as:

$$k(x, x'; t, \tau) = k_x(x, x') k_t(t, \tau) \quad (21)$$

Without loss of generality, we assume that the process of diffusion in space is uniform on statistics and stationary random process on time, so the spatial diffusion kernel function $k_x(x, x')$ is a function of $(x-x')$, and the diffusion kernel function $k_t(t, \tau)$ of time is a function of $(t-\tau)$.

If we assume $k_x(x, x')$, $k_t(t, \tau)$ with the property of negative power law,

$$k_x(x, x') = \frac{D}{\Gamma(-\alpha)} \frac{1}{(x-x')^{\alpha-1}}, \quad (22)$$

$$k_t(t, \tau) = \frac{1}{\Gamma(\gamma)} \frac{1}{(t-\tau)^{1-\lambda}} \quad (23)$$

In this expression, D , α , γ are constant, $\Gamma(-\alpha)$, $\Gamma(\gamma)$ are gamma function. Substituting the expression (19) and (20) to expression (18), then derivative the t and x of both side of equality sign in expression (18), it can get the expression like below:

$$J = D \frac{1}{\Gamma(\gamma)} \frac{\partial}{\partial t} \int_0^r \frac{1}{(t-\tau)^{1-\gamma}} d\tau \frac{1}{\Gamma(-\alpha)} \frac{\partial}{\partial x} \int_0^x \frac{1}{(x-x')^{\alpha-1}} C(x', \tau) dx' \quad (24)$$

Further, bring in the definition of the Riemann-Liouville fractional derivative, the result is like below:

$$J = {}_0^{RL} D_t^{1-\gamma} \left(D \frac{\partial^{\alpha-1}}{\partial x^{\alpha-1}} C(x, t) \right) \quad (25)$$

Then, we can get the diffusion equation with fractional order,

$$\frac{\partial C(x, t)}{\partial t} = {}_0^{RL} D_t^{1-\gamma} \left(D \frac{\partial^\alpha}{\partial x^\alpha} C(x, t) \right), 1 \leq \alpha \leq 2 \quad (26)$$

Comparing to the classical diffusion equation, the diffusion equation of fractional order has more implications: the property of generalized flux, it can describe the physical phenomenon more accurate.

In turn, we will study the basic set of equation of atmospheric dynamic with fractional order, although these equations already have no application in practice, however, we can apply one of which when simulation and gradually introduce, we may get some useful results.

4 The Cantor-Type of the Basic Set of Equation in Mesoscale Meteorological Model

4.1 CONSERVATION OF MASS WITH FRACTIONAL ORDER

In this section, we start with conversation of mass on Cantor Sets with fractional order. The mass of fractal fluid in $v^{(\nu)}$ is defined through [2].

$$M = \iiint_{v^{(\nu)}} \rho dv^{(\nu)}. \tag{27}$$

And we can get:

$$\frac{\partial^\alpha M}{\partial t^\alpha} = -\iint_{s^{(\beta)}} \rho v ds^{(\beta)}. \tag{28}$$

Using Divergence Theorem of local fractional field, we can get:

$$\begin{aligned} & \iiint_{v^{(\nu)}} \frac{\partial^\alpha}{\partial t^\alpha} \rho dv^{(\nu)} + \iint_{s^{(\beta)}} \rho v ds^{(\beta)} \\ &= \iiint_{v^{(\nu)}} \left[\frac{\partial^\alpha \rho}{\partial t^\alpha} + \nabla^\alpha(\rho v) \right] dv^{(\nu)} = 0 \end{aligned} \tag{29}$$

This implies that:

$$\frac{\partial^\alpha \rho}{\partial t^\alpha} + \nabla^\alpha(\rho v) = 0. \tag{30}$$

This is called continuity equation with fractional order on fractal materials, which also have the formulation (31):

$$\begin{aligned} \frac{\partial^\alpha \rho}{\partial t^\alpha} + \nabla^\alpha(\rho v) &= \frac{\partial^\alpha \rho}{\partial t^\alpha} + v(\nabla^\alpha \rho) + \rho(\nabla^\alpha v) \\ &= \frac{D^\alpha \rho}{Dt^\alpha} + \rho(\nabla^\alpha v) = 0 \end{aligned} \tag{31}$$

Where in the fractal material derivative of the fluid density ρ is expressed as:

$$\frac{D^\alpha \rho}{Dt^\alpha} = \frac{\partial^\alpha \rho}{\partial t^\alpha} + v(\nabla^\alpha \rho). \tag{32}$$

If the fractal fluid is incompressible, we also get:

$$\frac{\partial^\alpha \rho}{\partial t^\alpha} + v(\nabla^\alpha \rho) = 0 \Leftrightarrow \frac{D^\alpha \rho}{Dt^\alpha} = 0, \text{ or } \nabla^\alpha v = 0. \tag{33}$$

The corresponding form of the equation of conservation of mass in spherical coordination as below,

$$\begin{aligned} & \frac{d^\alpha \rho}{dt^\alpha} + \rho \left(\frac{1}{r^\alpha \cos^\alpha \varphi} \frac{\partial^\alpha u}{\partial \lambda^\alpha} + \frac{1}{r^\alpha} \frac{\partial^\alpha v}{\partial \varphi^\alpha} + \frac{\partial^\alpha w}{\partial r^\alpha} \right) \\ & + \rho \left(-\frac{v}{r^\alpha} \tan^\alpha \frac{\partial^\alpha w}{r^\alpha} \right) = 0 \end{aligned} \tag{33a}$$

Can simplify as,

$$\begin{aligned} & \frac{\partial^\alpha \rho}{\partial t^\alpha} + \frac{1}{r^\alpha \cos^\alpha \varphi} \frac{\partial^\alpha \rho u^\alpha}{\partial \lambda} + \frac{1}{r^\alpha} \frac{\partial^\alpha \rho v^\alpha}{\partial \varphi^\alpha} \\ & + \frac{\partial^\alpha \rho w^\alpha}{\partial r^\alpha} - \frac{\rho v^\alpha}{r^\alpha} \tan^2 \varphi + \frac{2\rho w^\alpha}{r^\alpha} = 0 \end{aligned} \tag{33b}$$

4.1.1 Remark

Similarly, we can get the expression like below:

(1) Conservation of heat with fractional order.

$$\frac{\partial^\alpha \rho}{\partial t^\alpha} + \nabla^\alpha(\rho v) = S_\theta. \tag{34}$$

In this expression, S_θ represent the source and sink of heat.

(2) Conservation of motion with fractional order.

$$\frac{\partial^\alpha V}{\partial t^\alpha} = -V \nabla^\alpha V - \frac{1}{\rho} \nabla^\alpha(\rho v) - gk - 2\Omega \times V. \tag{35}$$

(3) Conservation equation of energy with fractional order.

Similarly, fractional order energy conservation equation in fractal media can be express like below:

$$\begin{aligned} & \frac{d^\alpha(\rho h)}{dt^\alpha} + \frac{\partial^\alpha(\rho u h)}{\partial x^\alpha} + \frac{\partial^\alpha(\rho u h)}{\partial y^\alpha} + \frac{\partial^\alpha(\rho w h)}{\partial z^\alpha} \\ &= -P \text{div}^\alpha u + \text{div}^\alpha(k \Delta^\alpha T) + \Phi + s_n \end{aligned} \tag{36}$$

In the expression (36) above, k is the coefficient of thermal conductivity of fluid. s_n is the thermal source in fluid. Φ called dissipation coefficient is the part of which thermal energy transformed from mechanical energy. $P \text{div}^\alpha u$ is the force to work on the fluid.

4.2 VORTICITY AND VORTICITY AND EQUATION WITH FRACTIONAL ORDER

Vorticity, that is to describe the feature of tiny clumps in air rotative-field. Due to the Earth's rotation, we can often see that the vortex of motion often occur in the process of movement, such as cyclone, anticyclone, typhoon etc.

In non-uniform three dimensional flow field, air tiny clumps at the same time would rotate around the X, Y, Z axis, that is, there are three vorticity components: ζ, η, ξ ,

the sum of these three vector known as the speed of vorticity in three-dimensional space, We represent it as rotation.

Based on above, we can get the form of speed vorticity with fractional order:

$$\begin{aligned} rot^\alpha V &= \zeta e_1^\alpha + \eta e_2^\alpha + \xi e_3^\alpha = \left(\frac{\partial^\alpha w}{\partial y^\alpha} - \frac{\partial^\alpha u}{\partial z^\alpha} \right) e_1^\alpha \\ &+ \left(\frac{\partial^\alpha u}{\partial z^\alpha} - \frac{\partial^\alpha w}{\partial x^\alpha} \right) e_2^\alpha + \left(\frac{\partial^\alpha u}{\partial x^\alpha} - \frac{\partial^\alpha u}{\partial y^\alpha} \right) e_3^\alpha \end{aligned} \quad (37)$$

Take a further step; we can get the vortex equation with fractional order by using vector operation, which describe the vortex motion.

The vector form of equations of motion with fractional order like below:

$$\frac{d^\alpha \vec{V}}{dt^\alpha} = g - \frac{1}{\rho} \nabla^\alpha p - 2\Omega \times \vec{V} + F. \quad (38)$$

According to

$$\frac{d^\alpha V}{d^\alpha t} = \frac{\partial V}{\partial t} + \nabla^\alpha \left(\frac{V^2}{2} \right) + 2\Omega \times V. \quad (39)$$

Therefore, equation of motion (39) would rewrite as:

$$\frac{\partial^\alpha \vec{V}}{\partial t^\alpha} + \omega_0 \times \vec{V} = g - \frac{1}{\rho} \nabla^\alpha p - \nabla \left(\frac{\vec{V}^2}{2} \right) + F. \quad (40)$$

Do the rotation operation to (40), we can get:

$$\begin{aligned} \nabla^\alpha \times g &= \nabla^\alpha \times (\nabla^\alpha \phi) = 0, \\ \nabla^\alpha \times \left(-\frac{1}{\rho} \nabla^\alpha p \right) &= -\nabla^\alpha \frac{1}{\rho} \times \nabla^\alpha p = B. \end{aligned} \quad (41)$$

And using (42), (43),

$$\nabla^\alpha \times (A \times B) = (B \times \nabla^\alpha) A - (A \nabla^\alpha B) - B(\nabla^\alpha A), \quad (42)$$

$$\begin{aligned} \nabla^\alpha \times (\omega_a \times V) &= (V \nabla^\alpha) \omega_a - (\omega_a \nabla^\alpha) V + \omega_a (\nabla^\alpha V) \\ - V(\nabla^\alpha \omega_a) &= (V \nabla^\alpha) \omega_a - (\omega_a \nabla^\alpha) V + \omega_a (\nabla^\alpha V) \end{aligned} \quad (43)$$

And (44),

$$\nabla^\alpha \omega_a = \nabla(\omega + 2\pi) = \nabla^\alpha \omega + 2\nabla^\alpha \Omega = \nabla^\alpha (\nabla^\alpha \times V) = 0. \quad (44)$$

Therefore, we get the vorticity equation with fractional order as below:

$$\frac{d^\alpha \omega_a}{dt^\alpha} = -(\omega_a \nabla^\alpha) V + (\nabla^\alpha V) \omega_a = B + \nabla^\alpha \times F. \quad (45)$$

5 Conclusions

In the present work, we present some basic set of equations in mesoscale meteorological model on the Cantor sets derived from local fractional vector calculus. These could be applied to describe atmospherical flow. The later used fractional calculus is continuous and differential quantities as classical result; what's more, the latter is local fractional continuous and no differential quantities. The classical result is obtained in case of fractal space-time dimension, which is equal to (1).

Acknowledgements

This research was financially supported by the MWR public sector research and special funds - the most stringent in arid zone water resources management key technologies (201301103) and National Nature Science Foundation of China (NSFC) under Grant No. 41130641, 41201025 and The Ministry of Education Key Laboratory of Eco-Oasis Open Topic-Moisture change in Central Asia and its influence on precipitation in Xinjiang province (XJDX0201-2013-07) and the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant No:2012BAC23B01). This financial support is gratefully acknowledged and appreciated.

References

- [1] Podlubny I 1999 Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications *Mathematics in Science and Engineering* 198 Academic Press, San Diego, Cali, USA
- [2] Yang X J 2011 *Local Fractional Functional Analysis and Its Applications*, Asian Academic publisher Limited, Hong Kong, China
- [3] Yang X J 2011 Local fractional integral transforms *Progress in Nonlinear Science* 10(4) 1-25
- [4] Chang Fu-Xuan, Chen Jin, Huang Wei 2006 Anomalous diffusion and fractional advection-diffusion equation *Journal of statistical Physics* 123(1) 89-100
- [5] Chen W, Sun H, Zhang X, Korosak D 2010 Anomalous diffusion modelling by fractal and fractional derivatives *Computers & Mathematics with Applications* 59(5) 1754-58
- [6] Kolwankar K M, Gangal A D 1998 Local fractional Fokker-Planck equation *Physical Review Letters* 80(2) 214-7
- [7] Carpinteri A, Chiaia B, Cornetti P 2001 Static-kinematic duality and the principle of virtual work in the mechanics of fractal media *Computer Methods in Applied Mechanics and Engineering* 191(1-2) 3-19
- [8] Adda F B, Cresson J 2001 about non-differentiable functions *Journal of Mathematical Analysis and Applications* 263(2) 721-37
- [9] Tarasov V E 2005 Continuous medium model for fractal media *Physics Letters, Section A* 336(2-3) 167-74
- [10] Yang X J 2013 Cylindrical-coordinate method for differential equations with local fractional derivatives *Physics Letters, Section A* 377 (28-30) 1696-700

[11] Yang X J, Baleanu D 2013 Fractal heat conduction problem solved by local fractional variation iteration method *Thermal Science* 17(2) 625-8

[12] Hu M-S, Agarwal R P, Yang X -J 2012 Local fractional Fourier series with application to wave equation in fractal vibrating string *Abstract and Applied Analysis* p 15 Article ID 567401

[13] Drapaca C S, Sivaloganathan S 2012 A fractional model of continuum mechanics *Journal of Elasticity* 107(2) 105-23

[14] Tarasov V E 2005 Fractional hydrodynamic equations for fractal media *Annals of Physics* 318(2) 286-307

[15] Babakhani A, Daftardar-Gejji V 2002 On calculus of local fractional derivatives *Journal of Mathematical Analysis and Applications* 270(1) 66 -79

Meng Xian-Yong, Liu Yang, Gao Yu-Xiao, Wang Mei, Liu Zhi-Hui

[16] Jumarie G 2009 Table of some basic fractional calculus formulae derived from a modified Riemann-Liouville derivative for non differentiable functions *Applied Mathematics Letters* 22(3) 378-85

[17] Chen W, Sun H, Zhang X, Korosak D 2010 Anomalous diffusion modelling by fractal and fractional derivatives *Computers & Mathematics with Applications* 59(5) 1754-8

[18] Jumarie G 2009 Table of some basic fractional calculus formulae derived from a modified Riemann-Liouville derivative for non differentiable functions *Applied Mathematics Letters* 22(3) 378-85

[19] Zaslavsky G M 2002 Chaos, fractional kinetics, and anomalous transport *Physics Reports* 371(6) 461-580

[20] Yang Y J, Baleanu D, Yang X J 2013 A local fractional variational iteration method for Laplace equation within local fractional operators *Abstract and Applied Analysis* 1-6

Authors	
	<p>Meng, Xian-Yong, born in December, 1987, Urumqi City, Xinjiang Province, P.R. China</p> <p>Current position, grades: doctoral degree at Xinjiang Institute of ecology and geography, China Academy of Sciences and Xinjiang University.</p> <p>Scientific interest: Hydrology and water resources, the integration of 3S technology and Simulation of land surface process.</p> <p>Publications: Published 27 academic papers (include 13 SCI/EI cite papers) in various journals.</p> <p>Experience: who win the first prize of the 2012 Xinjiang water conservancy industry outstanding young scientific papers, three Software patent. Dr. Meng, Xian-Yong is the teacher of the education Centre of Institute of Computing Technology, Chinese Academy of Science.</p> <p>Major in: Water resource background in systems analysis, including water resource planning, modelling, system optimization and simulation under uncertainty, impact assessment, environmental database development, and environmental management. Strong skill in Building numerical model by using SWAT (Soil and water assessment tool) models.</p>
	<p>Liu, Yang, born in February, 1987, Urumqi City, Xinjiang Province, P.R. China</p> <p>Current position, grades: doctoral degree at Xinjiang University.</p> <p>Scientific interest: Simulation of land surface process.</p> <p>Publications: Published 5 academic papers (include 4 SCI/EI cite papers) in various journals.</p>
	<p>Gao, Yu-Xiao, born in July, 1987, Urumqi City, Xinjiang Province, P.R. China</p> <p>Current position, grades: doctoral degree at Xinjiang Institute of ecology and geography, China Academy of Sciences and Xinjiang University.</p> <p>Scientific interest: Hydrology and water resources. LUCC</p> <p>Publications: Published 1 academic papers (include 1 EI cite papers).</p>
	<p>Wang, Mei, born in August, 1990, Urumqi City, Xinjiang Province, P.R. China</p> <p>Current position, grades: master degree at Xinjiang University.</p> <p>Scientific interest: The integration of 3S technology.</p>
	<p>Liu, Zhi-Hui, born in July, 1957, Urumqi City, Xinjiang Province, P.R. China</p> <p>Current position, grades: Ph.D., doctoral tutor, director of the Institute of ecological environment in arid Xinjiang University, and deputy director of the Ministry of Education Key Laboratory of Oasis Ecology. It also serves as: 1. Xinjiang Geographic Society vice president, executive director; 2. Xinjiang Institute of Soil and Water Conservation, vice chairman, executive director; 3. Xinjiang Remote Sensing Application Technology Association executive director; 4. Xinjiang Institute of Water Conservancy executive director of the Academic Committee, deputy director of the professional Committee of hydrology.</p> <p>Scientific interest: Hydrology and water resources, the integration of 3S technology and Simulation of land surface process.</p> <p>Publications: Published 170 academic papers (include 30 SCI/EI cite papers) in various journals.</p> <p>Experience: has long been engaged in teaching and research of water resources and the environment, ecology and environmental planning, land resources in areas such as planning and evaluation, who had studied in Austria and Switzerland, and distributed watershed model, "3S" technologies, decision support systems, water management other aspects of the new research results.</p>

DSM control strategy of solid oxide fuel cell distributed generation system

Junsheng Jiao*

School of Electrical Engineering, Tongling University, Tongling City, Anhui Province, China, 244061

Received 6 October 2013, www.tsi.lv

Abstract

The distributed power generation system of fuel cell has received the widespread attention in recent years for its pro-environment and high efficiency. The solid oxide fuel cell (SOFC) is a new power system, which turns chemical energy of fuels into electrical energy directly at middle and high temperature. Mathematical model of SOFC is analysed and the main circuit structure of independent power generation system is designed in this paper. The traditional PI control is adopted in DC/DC boost circuit to ensure the fuel cell can provide a stable DC output voltage. The three-phase inverter is used in DC/AC circuit and the dynamic equation derived from three-phase inverter is transformed to the synchronous rotating coordinate system. An innovative discrete sliding mode (DSM) control technology is applied and the system output stable sinusoidal AC voltage. The simulation experiments prove that SOFC power systems, in a certain range, can quickly and dynamically trace the change of the input voltage and load using DSM control.

Keywords: SOFC, Discrete Sliding Mode, Control, DC/AC, Converter

1 Introduction

As the global conventional energy is drying up and the environmental pollution problems are increasingly more serious, the distributed generation, in the form of small scale, scattered layout near the users and output the power independently, has received extensive attention for its advantages of pro-environment, high efficiency and flexible usage [1-2]. It will be an important part of micro grid in the future and a kind of competitive power in future power market.

Among various kinds of distributed generation system, fuel cell, wind, water, solar and other renewable energy have received great attention. The fuel cell is not like the wind and hydropower limited by geographical factors. It has a constant power output and advantages of high energy conversion efficiency, clean, no pollution and low noise [3-4]. It is suitable for either centralized power supply or distributed power supply and it also can be used as power supply of space vehicle, transportation. So it possesses more special application value.

Solid oxide fuel cell (SOFC) belongs to the third generation of fuel cell technology, the integrated thermoelectric efficiency of independent SOFC power system is 46% when the output power is 109KW [5-8]. It has the advantages of wide fuel adaptability, no leakage, high efficiency of comprehensive utilization and long service life. Therefore, it arouses people attention and is a hot and key point in the research of fuel cell. SOFC can be used as distributed power of network for fixed power generation. It also can provide mobile power supply for

transportation tools, such as ships, vehicles, etc. [9-11]. Therefore, it is significant to develop the SOFC technology for improving the energy structure and maximizing the use of the existing energy and rare earth resources.

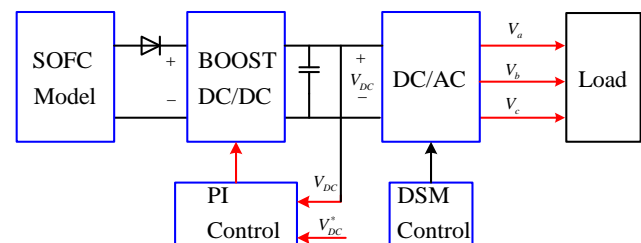


FIGURE 1 Main circuit of SOFC generation

Due to the output power of fuel cell is limited by its physical properties, when the load is changes, the response speed is slow. The SOFC power generation system designed in this paper uses DC/DC power converter to provide stable DC voltage for the use of three-phase DC/AC power converter and the discrete sliding mode control is used in DC/AC circuit. The simulation experimental results show that, with the support of the power electronic converter equipment, SOFC can provide stable power output for the network and fast load tracking capability.

* Corresponding author - Tel: +86-0562-588-2089 fax: +86-0562-588-2089; E-mail: jiaotlu@sina.com

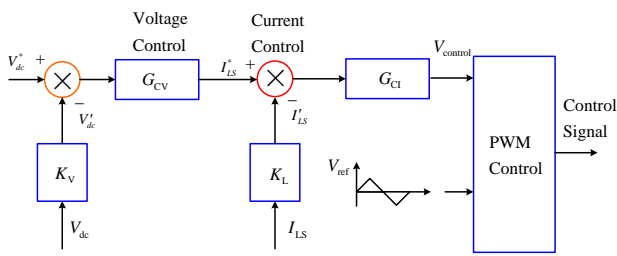
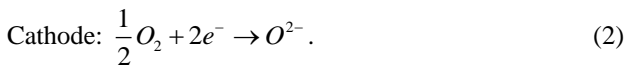
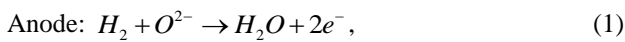


FIGURE 2 DC/DC PI control

2 SOFC model

The basic electrode reaction in the SOFC is different from that of other kinds of fuel cell [12]. The reaction is explained by the following formula:



At the anode, hydrogen gas reacts with ions to create water in gaseous state and electrical energy is discharged in the configuration of the electrons. At the cathode, oxygen reacts with the electrons taken from the electrode, and oxygen ion is engendered. To count the open-circuit electromotive force of a stack of fuel cells in series, refer to the well-known Nernst formula [12]. In the formula above, E can be calculated by Nernst equation:

$$E = N_0 E_0 + E_f \ln\left(\frac{p_{H_2} p_{O_2}^{0.5}}{p_{H_2O}}\right), \quad (3)$$

where E_0 is the voltage associated with the reaction-free energy of a cell, $E_f = N_0 RT / 2F$, R is the gas constant (8.31J/mol.K), T is the SOFC operating temperature, F is the Faraday constant (96487C/mol), and p_i are the partial pressures of hydrogen, oxygen and water.

3 Main circuit structure

Electrical energy from the fuel cell itself is usually not stable and the output voltage tends to be weak, it cannot use directly for load or electrical power grid. In order to meet the load demand, some corresponding converter

devices is required. Main circuit of SOFC generation is in Figure 1. The first stage is usually DC/DC processing circuit. The converter turns the unstable output voltage of fuel cell into input voltage V_{DC} of the back stage inverter DC bus. The second stage is the inverter control, it is a core part of the distributed generation system, the current hot research spot and the technical difficulties are mainly concentrated in this part. The output DC voltage of SOFC boosts 35V by the boost converter; then three-phase AC sine voltage for load can be obtained from three-phase DC/AC inverter; after LC filter circuit 220V/50Hz AC output voltage is obtained.

4 DC/DC converter control

DC/DC boost converter control chart is exposed in Figure 2, V_{DC}^* is set as the boost instruction voltage. If the voltage is not reach the set value, the required current command produced by the voltage controller G_{CV} is compared with actual inductance current of converter, then the control command signal generated by the current controller is compared with the triangle wave to produce pulse width modulation signal, the duty ratio of the power switch is adjusted by control signals, the goal is to change the output voltage and achieve the stable output voltage.

The traditional PI control is illustrated in Figure 2. According to the classic laws of Ziegler -Nichols, V_{DC} is 35V. K_V of the inner current loop controller in DC/DC converter can be obtained as 19.72. In outer circle control loop, that is the voltage control loop and K_L is 82.36.

5 Three-phase DC/AC converter

Three-phase and three-leg DC/AC power converter, the structure is illustrated in Figure 3. DC shunt capacitance on the side stabilizes the output voltage. Three-phase voltage is generated by the on-off of control switch. Three legs and six switch components can be seen in Figure 3, two switches constitute a leg, each leg consists of two power switch element and their states of connection or break are complementary. That is when the upper leg switch VT_1 , VT_3 and VT_5 are on, the lower leg switch VT_4 , VT_6 and VT_2 are off, and vice versa. If the upper leg switch and lower leg switch of the same leg turn on, it will cause instantaneous short circuit and produce large current to burn down switch components.

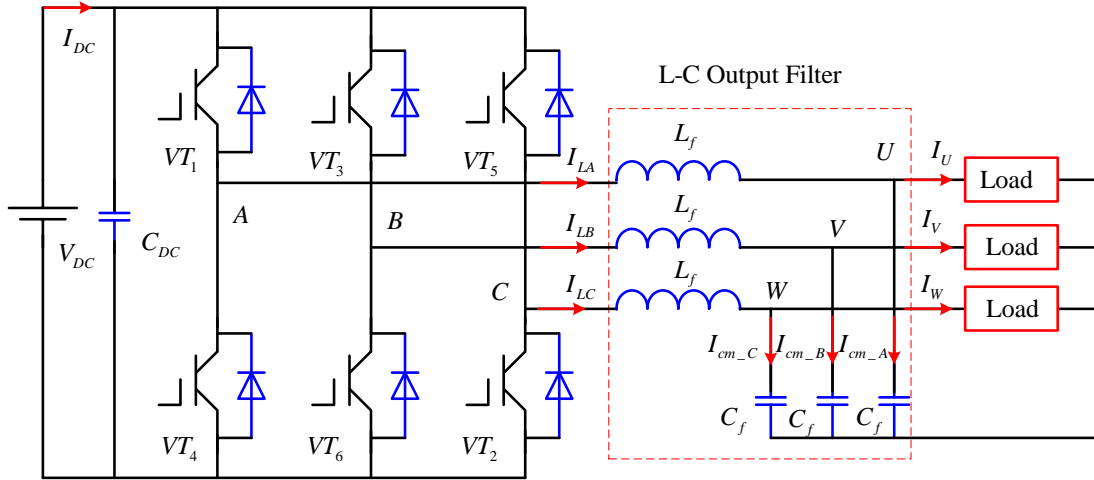


FIGURE 3 Three-phase DC/AC circuit

The circuit model explained by Figure 3 makes use of the following quantities [13]. The inverter output line-to-line voltage is expressed as the vector $V_i=[V_{iAB} \ V_{iBC} \ V_{iCA}]^T$, and the three-phase inverter output currents are I_{LA}, I_{LB} and I_{LC} . According as these currents, a vector is defined as $I_i=[I_{iAB} \ I_{iBC} \ I_{iCA}]^T=[I_{iA}-I_{iB} \ I_{iB}-I_{iC} \ I_{iC}-I_{iA}]^T$. Also, the line to line load voltage and phase load current vectors can be represented by $V_L=[V_{LAB} \ V_{LBC} \ V_{LCA}]^T$ and $I_L=[I_{LA} \ I_{LB} \ I_{LC}]^T$ respectively. On the L-C output filter, the following current and voltage equations are obtained after elementary calculation.

Voltage equations:

$$\begin{cases} \frac{dV_{LAB}}{dt} = \frac{1}{3C_f} I_{iAB} - \frac{1}{3C_f} (I_{LA} - I_{LB}) \\ \frac{dV_{LBC}}{dt} = \frac{1}{3C_f} I_{iBC} - \frac{1}{3C_f} (I_{LB} - I_{LC}) \\ \frac{dV_{LCA}}{dt} = \frac{1}{3C_f} I_{iCA} - \frac{1}{3C_f} (I_{LC} - I_{LA}) \end{cases} \quad (4)$$

Current equations:

$$\begin{cases} \frac{dI_{iAB}}{dt} = -\frac{1}{L_f} V_{LAB} + \frac{1}{L_f} V_{iAB} \\ \frac{dI_{iBC}}{dt} = -\frac{1}{L_f} V_{LBC} + \frac{1}{L_f} V_{iBC} \\ \frac{dI_{iCA}}{dt} = -\frac{1}{L_f} V_{LCA} + \frac{1}{L_f} V_{iCA} \end{cases} \quad (5)$$

Rewrite equation (4) and equation (5) into a vector form, respectively:

$$\begin{cases} \frac{dV_L}{dt} = \frac{1}{3C_f} I_i - \frac{1}{3C_f} T_i I_L \\ \frac{dI_i}{dt} = -\frac{1}{L_f} V_L + \frac{1}{L_f} V_i \end{cases}, \quad (6)$$

where $T_i = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$.

6 Synchronous rotating transformation

The influence of mutual coupling and time-varying exists in the dynamic equations of three-phase coordinate system and the design of the controller will be more complex.

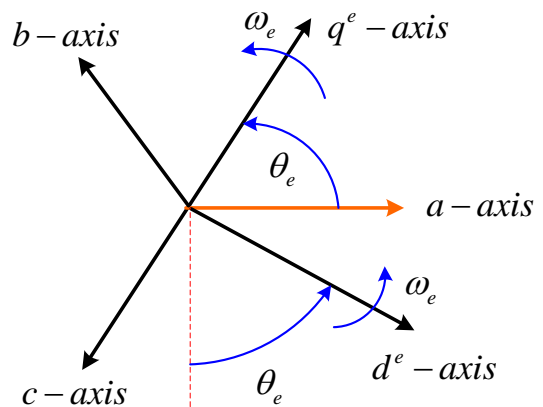


FIGURE 4 Synchronous rotating coordinate transformation

The derived dynamic equations can be converted into the synchronous rotating coordinate system before the design of controller. The geometric relations between three-phase coordinate system and the synchronous rotating coordinate system are shown as Figure 4.

The conversion relation is:

$$\begin{bmatrix} f_q^e \\ f_d^e \\ f_0^e \end{bmatrix} = T_{qd0}^e \begin{bmatrix} f_a \\ f_b \\ f_c \end{bmatrix} = \frac{2}{3} \begin{bmatrix} \cos \theta_e & \cos(\theta_e - \frac{2\pi}{3}) & \cos(\theta_e + \frac{2\pi}{3}) \\ \sin \theta_e & \sin(\theta_e - \frac{2\pi}{3}) & \sin(\theta_e + \frac{2\pi}{3}) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} f_a \\ f_b \\ f_c \end{bmatrix} \quad (7)$$

The inverse transform of the equation (7) is:

$$\begin{bmatrix} f_a \\ f_b \\ f_c \end{bmatrix} = [T_{qd0}^e]^{-1} \begin{bmatrix} f_q^e \\ f_d^e \\ f_0^e \end{bmatrix} = \begin{bmatrix} \cos \theta_e & \sin \theta_e & 1 \\ \cos(\theta_e - \frac{2\pi}{3}) & \sin(\theta_e - \frac{2\pi}{3}) & 1 \\ \cos(\theta_e + \frac{2\pi}{3}) & \sin(\theta_e + \frac{2\pi}{3}) & 1 \end{bmatrix} \begin{bmatrix} f_q^e \\ f_d^e \\ f_0^e \end{bmatrix} \quad (8)$$

where f denotes either a voltage or a current variable; $\theta_e = \int 2\pi f_e dt$; $f_e = 50\text{Hz}$.

Coordinating the transformation formula, the equation (4) and equation (5) are converted into the voltage and current equation of quadrature axis and direct axis in synchronous rotating coordinates, as shown in the equation (9) and equation (10) [14-15].

$$\begin{bmatrix} \frac{d}{dt} V_{q_UVM}^e \\ \frac{d}{dt} V_{d_UVM}^e \end{bmatrix} = \begin{bmatrix} 0 & -\omega_e \\ \omega_e & 0 \end{bmatrix} \begin{bmatrix} V_{q_UVM}^e \\ V_{d_UVM}^e \end{bmatrix} + \begin{bmatrix} \frac{1}{C_f} & 0 \\ 0 & \frac{1}{C_f} \end{bmatrix} \begin{bmatrix} I_{q_abc}^e \\ I_{d_abc}^e \end{bmatrix} + \begin{bmatrix} -\frac{1}{C_f} & 0 \\ 0 & -\frac{1}{C_f} \end{bmatrix} \begin{bmatrix} I_{q_UVM}^e \\ I_{d_UVM}^e \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} \frac{d}{dt} I_{q_UVM}^e \\ \frac{d}{dt} I_{d_UVM}^e \end{bmatrix} = \begin{bmatrix} 0 & -\omega_e \\ \omega_e & 0 \end{bmatrix} \begin{bmatrix} I_{q_abc}^e \\ I_{d_abc}^e \end{bmatrix} + \begin{bmatrix} \frac{1}{L_f} & 0 \\ 0 & \frac{1}{L_f} \end{bmatrix} \begin{bmatrix} V_{q_abc}^e \\ V_{d_abc}^e \end{bmatrix} + \begin{bmatrix} -\frac{1}{L_f} & 0 \\ 0 & -\frac{1}{L_f} \end{bmatrix} \begin{bmatrix} V_{q_UVM}^e \\ V_{d_UVM}^e \end{bmatrix} \quad (10)$$

where $V_{q_abc}^e, V_{d_abc}^e$ is the output voltage of quadrature axis and direct axis for three phase DC/AC inverter respectively; $V_{q_UVM}^e, V_{d_UVM}^e$ is load voltage of quadrature axis and direct axis for three phase DC/AC inverter respectively; $I_{q_abc}^e, I_{d_abc}^e$ is the output current of quadrature axis and direct axis for three phase DC/AC inverter respectively; $I_{q_UVM}^e, I_{d_UVM}^e$ is the load current of quadrature axis and direct axis for three phase DC/AC inverter respectively.

7 Discrete sliding mode control

The design of discrete sliding mode controller includes two aspects [16]: one is to seek the sliding mode plane function, which makes the system sliding mode motion on the surface stable gradually and has good quality. The second is the designing of variable structure control, which can make the system reach the sliding surface from any point of phase space in finite time and form sliding mode control at sliding surface.

The continuous-time state space equation (9) and equation (10) of the plant system can be expressed to include dynamics of DSM below:

$$\begin{cases} \dot{X}(t) = AX(t) + Bu(t) + Ed(t) \\ y(t) = CX(t) \end{cases} \quad (11)$$

where $X(t)$ is the system state vector, $X(t) \in R^n$; $u(t)$ is the system control vector, $u(t) \in R^m$; $d(t)$ is the system matching noise, $d(t) \in R^m$.

Given the sampling period T_z , the equation (11) can be transformed to the following discrete-time state space equation:

$$\begin{cases} X(k+1) = A^*X(k) + B^*u(k) + E^*d(k) \\ y(k) = CX(k) \end{cases} \quad (12)$$

where $A^* = e^{AT_z}$, $B^* = \int_0^{T_z} e^{A(T_z-\tau)} B d\tau$,

$E^* = \int_0^{T_z} e^{A(T_z-\tau)} E d\tau$.

In order to control the output $y(k)$ to follow the reference $y_{ref}(k)$, a sliding mode manifold may be selected in the following.

$$s(k) = y(k) - y_{ref}(k) = CX(k) - y_{ref}(k), \quad (13)$$

where $y_{ref}(k) = I_{cmd_iqd}(k)$.

In other words, when the discrete sliding mode exists, which means $s(k)=0$, the output $y(k)$ is identical to the reference $y_{ref}(k)$. Therefore, the discrete sliding mode exists if the control input $u(k)$ is designed as following:

$$\begin{aligned} s(k+1) &= y(k+1) - y_{ref}(k+1) \\ &= CA^*X(k) + CB^*u(k) + CE^*d(k) - y_{ref}(k+1) = 0 \end{aligned} \quad (14)$$

The control algorithm that satisfies equation (14) and yields motion in the manifold $s(k)=0$ is called equivalent control. For the given system, the equivalent control $u_{eq}(k)$ is given as follows:

$$u_{eq}(k) = (CB^*)^{-1} [y_{ref}(k+1) - CA^*X(k) - CE^*d(k)] \quad (15)$$

Then the required control command can be calculated.

Discrete slide model control is based on the double loop control of voltage and current and turns the voltage equation (9) into the discrete state equation [17-18].

$$\begin{aligned} \begin{bmatrix} V_{q_UVW}^e(k+1) \\ V_{d_UVW}^e(k+1) \end{bmatrix} &= A_1^* \begin{bmatrix} V_{q_UVW}^e(k) \\ V_{d_UVW}^e(k) \end{bmatrix} + B_1^* \begin{bmatrix} i_{q_abc}^e(k) \\ i_{d_abc}^e(k) \end{bmatrix}, \\ &+ D_1^* \begin{bmatrix} i_{q_UVW}^e(k) \\ i_{d_UVW}^e(k) \end{bmatrix} \end{aligned} \quad (16)$$

where $A_1^* = \begin{bmatrix} 0 & -\omega_e \\ \omega_e & 0 \end{bmatrix}$, $B_1^* = \begin{bmatrix} \frac{1}{C_f} & 0 \\ 0 & \frac{1}{C_f} \end{bmatrix}$,

$$D_1^* = \begin{bmatrix} -\frac{1}{C_f} & 0 \\ 0 & -\frac{1}{C_f} \end{bmatrix}.$$

The feedback values of three-phase voltage and current are substituted into the equation (17) through synchronous rotating coordinate conversion. The filtering inductance current command $i_{q_abc_cmd}^*$, $i_{d_abc_cmd}^*$ can be acquired.

$$\begin{aligned} \begin{bmatrix} i_{q_abc_cmd}^e \\ i_{d_abc_cmd}^e \end{bmatrix} &= \\ B' \left\{ \begin{bmatrix} V_{q_UVW}^e \\ V_{d_UVW}^e \end{bmatrix} - A_1' \begin{bmatrix} V_{q_UVW}^e \\ V_{d_UVW}^e \end{bmatrix} - D_1' \begin{bmatrix} i_{q_UVW}^e \\ i_{d_UVW}^e \end{bmatrix} \right\}, \end{aligned} \quad (17)$$

where $A_1' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} A_1^*$, $B_1' = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} B_1^* \right\}^{-1}$,
 $D_1' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} D_1^*$.

The current loop equation (10) can be converted into discrete state equations.

$$\begin{aligned} \begin{bmatrix} i_{q_abc}^e(k+1) \\ i_{d_abc}^e(k+1) \end{bmatrix} &= A_2^* \begin{bmatrix} i_{q_abc}^e(k) \\ i_{d_abc}^e(k) \end{bmatrix} \\ &+ B_2^* \begin{bmatrix} V_{q_abc}^e(k) \\ V_{d_abc}^e(k) \end{bmatrix} + D_2^* \begin{bmatrix} V_{q_UVW}^e(k) \\ V_{d_UVW}^e(k) \end{bmatrix}, \end{aligned} \quad (18)$$

where $A_2^* = \begin{bmatrix} 0 & -\omega_e \\ \omega_e & 0 \end{bmatrix}$, $B_2^* = \begin{bmatrix} \frac{1}{L_f} & 0 \\ 0 & \frac{1}{L_f} \end{bmatrix}$,

$$D_2^* = \begin{bmatrix} -\frac{1}{L_f} & 0 \\ 0 & -\frac{1}{L_f} \end{bmatrix}.$$

Taking the filtering inductance current commands and three-phase voltage obtained by synchronous rotating coordinate transformation into equation (19), the output voltage adjustment of three-phase DC/AC converter can be obtained.

$$\begin{aligned} \begin{bmatrix} V_{q_abc_cmd}^e \\ V_{d_abc_cmd}^e \end{bmatrix} &= \\ B_2' \left\{ \begin{bmatrix} I_{q_abc}^e \\ I_{d_abc}^e \end{bmatrix} - A_2' \begin{bmatrix} I_{q_abc}^e \\ I_{d_abc}^e \end{bmatrix} - D_2' \begin{bmatrix} V_{q_UVW}^e \\ V_{d_UVW}^e \end{bmatrix} \right\}, \end{aligned} \quad (19)$$

where $A_2' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} A_2^*$, $B_2' = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} B_2^* \right\}^{-1}$,
 $D_2' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} D_2^*$.

8 Simulation results

The software of Matlab/Simulink is used for simulation analysis in the paper and Figure 5 shows the output voltage waveform of DC/DC converter. When the input voltage of SOFC drops down at $t=0.1s$, the change of the boost converter can be seen in Figure 5. As the decline of input voltage, the output voltage of boost converter drops down instantaneously. However, the output DC voltage can restore 35V at short time for the error compensation action of controller.

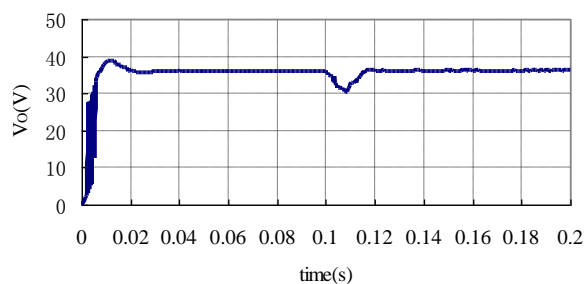


FIGURE 5 Output response of DC/DC

The independent operation mode is shown in Figure 6. The three-phase equilibrium resistive load, when the load power is 2000W, the output voltage and current of three-phase power converter are controlled by DSM controller. It can be seen from Figure 6, the peak value of output a phase voltage is about 311V and a peak value of phase current is about 4.28A.

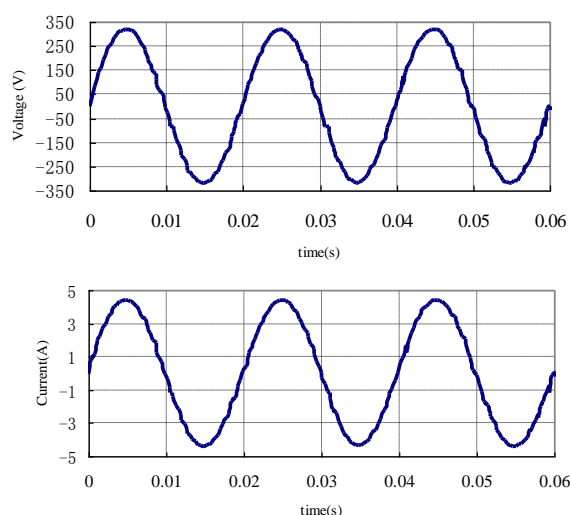


FIGURE 6 Output response of voltage and current of three-phase DC/AC inverter

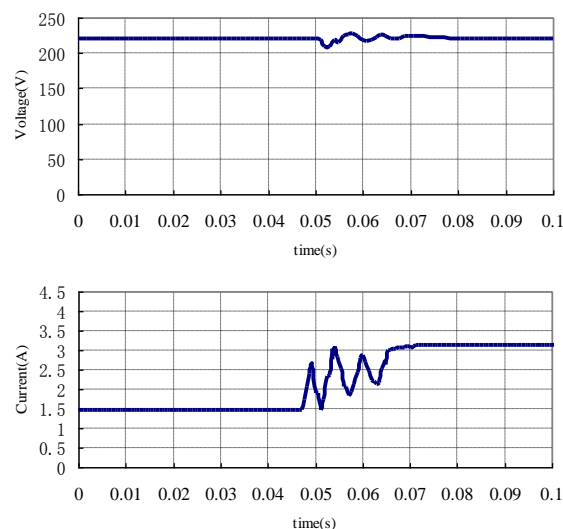


FIGURE 7 Output response of voltage and current RMS value for three-phase DC/AC inverter in load variation

When load is changed, Output response of voltage and current RMS value is shown in Figure 7 for three-phase DC/AC power converter. When it is increased from 990W to 2000W at 0.01s, the RMS (root mean square) value of voltage and current can be obtained. It can be seen from Figure 7, which the voltage can restore stable in 0.015s by DSM controller, the current can restore stable in 0.023s.

9 Conclusion

The SOFC generation system based on PI and discrete slide mode control is designed in this paper. Its output power can reach 2000W and it can be applied in the SOFC generation system. When the load of three-phase power converter is changed, the voltage and current can be effectively controlled by DSM controller with fast dynamic response. The convergence rate is fast and there is no overshoot. The output of power is stable. It is more suitable for distributed generation system at steady state. The simulation results verify the feasibility of control algorithm.

References

- [1] Sopian K, Wan Daud W 2006 Challenges and future developments in proton exchange membrane fuel cells *Renewable Energy* **31**(5) 719-27
- [2] Olsen F A, Berenguer M G, Molina 2010 Design of improved fuel cell controller for distributed generation systems *International Journal of Hydrogen Energy* **35**(11) 5974-80
- [3] Bauen A, Hart D, Chase A 2003 Fuel cells for distributed generation in developing countries-an analysis *International Journal of Hydrogen Energy* **28**(7) 695-701
- [4] Vijay P, Samantaray A K, Mukherjee A 2009 A bond graph model-based evaluation of a control scheme to improve the dynamic performance of a solid oxide fuel cell *Mechatronics* **19**(4) 489-502
- [5] Kaneko T, Brouwer J, Samuelsen G S 2006 Power and temperature control of fluctuating biomass gas fueled solid oxide fuel cell and

- micro gas turbine hybrid system *Journal of Power Sources* **160**(1) 316-25
- [6] Allag T, Das T 2012 Robust control of solid oxide fuel cell ultracapacitor hybrid system *IEEE Transactions on Control Systems Technology* **20**(1) 1-10
- [7] Li Y H, Rajakaruna S, Choi S S 2007 Control of a solid oxide fuel cell power plant in a grid-connected system *IEEE Transactions on Energy Conversion* **22**(2) 405-13
- [8] Komatsu Y, Kimijima S, Szmyd J S 2013 Numerical analysis on dynamic behavior of solid oxide fuel cell with power output control scheme *Journal of Power Sources* **233**(2) 232-45
- [9] Caisheng Wang, Nehrir M H 2007 Short-time overloading capability and distributed generation applications of solid oxide fuel cells *IEEE Transactions on Energy Conversion* **22**(4) 898-906
- [10] Du W, Wang H F, Zhang X F, Xiao L Y 2012 Effect of grid-connected solid oxide fuel cell power generation on power systems small-signal stability *IET Renewable Power Generation* **6**(1) 24-37
- [11] Saha A K, Chowdhury S, Chowdhury S P, Song Y H 2007 Application of solid-oxide fuel cell in distributed power generation *IET Renewable Power Generation* **1**(4) 193-202
- [12] Vikram M, Vinod M J, Steffen T, Olaf D 2012 A novel approach to model the transient behavior of solid-oxide fuel cell stacks *Journal of Power Sources* **214**(9) 227-38
- [13] Ito Y, Kawauchi S 1995 Microprocessor based robust digital control for UPS with three-phase PWM inverter *IEEE Transactions on Power Electronics* **10**(2) 196-204
- [14] Ilyas E 2006 Sliding mode control with PID sliding surface and experimental application to an electromechanical plant *ISA Transactions* **45**(1) 109-18
- [15] Castaneda C E, Loukianov A G, Sanchez E N, Castillo-Toledo B 2012 Discrete-time neural sliding-mode block control for a DC motor with controlled flux *IEEE Transactions on Industrial Electronics* **59**(2) 1194-1207
- [16] Schirone L, Celani F, Macellari M 2012 Discrete-time control for DC-AC converters based on sliding mode design *IET Power Electronics* **5**(6) 833-40
- [17] Hassan S, Seyyed M M K, Gholamreza V, Aria A 2009 Stabilizing unstable fixed points of discrete chaotic systems via quasi-sliding mode method *Communications in Nonlinear Science and Numerical Simulation* **14**(3) 839-49
- [18] Utkin V, Guldner J, Shi J 1999 *Sliding Mode Control in Electromechanical Systems* CRC Press: Boca Raton

Authors



Junsheng Jiao, born in October, 1968, Wuwei County, Anhui Province, P.R. China

Current position, grades: the Associate Professor of School of Electrical Engineering, Tongling University, China.

University studies: received his B.Sc. in Electrical Engineering and Automation from Anhui University of Science and Technology in China. He received his M.Sc. from Jiangsu University in China.

Scientific interest: His research interest fields include power electronics, renewable energy and simulation.

Publications: more than 10 papers published in various journals.

Experience: He has teaching experience of eight years, has completed three scientific research projects.

Micro-analysis of sub-ballast direct shearing under normal stress

Jing Guoqing^{1*}, Liu Guixian¹, Xu Yang¹, Zhang Jiong²

¹ Civil Engineering School, Beijing JiaoTong University, Beijing, P.R. China, 100044

² School of Civil Engineering, Shandong University, Jinan, P.R. China, 250100

Received 27 January 2014, www.tsi.lv

Abstract

This paper presents sub-ballast material analysis through Discrete Element Modelling (DEM) method to investigate its direct shear tests behaviour. Laboratory direct shear box tests were conducted on granite sub-ballast aggregate samples under different normal stress. The size, particle size distribution (PSD) properties of the sub-ballast particles were considered in the analysis. Simultaneously, direct shear box DEM simulations were conducted for different normal stress conditions, the sub-ballast micro-characteristics of the direct shearing tests were analysed, the contact force, displacement etc. index variation interrelated with the direct shearing tests conditions. The shear stress-shear displacement curves predicted from the DEM simulations were in reasonably good agreement with laboratory test results at all normal stress levels that shear box tests were conducted at. The results of discrete element modelling and tests of direct shear test for sub-ballast were presented. Such analysis may provide useful information on the understanding of sub-ballast with discrete element models. The micro-analysis and the direct shear tests were of importance to investigate the sub-ballast behaviour under the track system.

Keywords: direct shearing test; DEM; sub-ballast, normal stress

1 Introduction

Sub-ballast is a free draining granular material that helps to distribute an induced cyclic load to the underlying subgrade at a reduced and acceptable level of stress, and filter the water [1]. The sub-ballast used in most newly built railway system is intended to prevent the mutual penetration or intermixing of the subgrade and the ballast and to reduce frost penetration. The sub-ballast granular is similar to highway base material, which is fine grained-smaller than ballast and better size particles, and it was compacted tight and dense with lower voids. It is imperative and necessary that the magnitude and distribution of stresses and displacements of sub-ballast under the track should be evaluated and investigated accurately. Large-scale biaxial and triaxial tests have been carried out to study the mechanical behaviour of sub-ballast materials under static and cyclic loading. Most works have been performed on ballast [2-4]. Large-scale laboratory tests such as direct shear have been used for studying the behaviour of railway sub-ballast materials, but microscopic response study receives less attention.

This paper focuses on the microscopic analysis of the sub-ballast shear tests. For this purpose, several tests were conducted on dry, clean granite sub-ballast of high speed lines using a medium-scale direct shear apparatus. In the tests, the materials were compacted dry and

subjected to normal stress and shear loading. Sub-ballast direct shear tests were conducted under 4 different normal stress levels. The results of tests were compared with the analysis of discrete element method (DEM) simulations. The paper outlines the simulations and presents tests results, overall mechanical load-deformation response. Quantitative analysis of the sub-ballast fabric using the contact normal force and distributions allowed observations on the fundamental mechanisms under the observed response. The results indicate that the normal loading influences the inherent fabric characteristics of sub-ballast.

2 Material and method

2.1 APPARATUS

A middle-scale direct shear apparatus (30×30×20 cm) was used for tests as illustrated in Figure 1 [5]. The apparatus has an upper and a lower shear box, and the sample was sheared strain-controlled by pushing the lower shear box horizontally. Two gauges were used to measure lateral displacements and shearing forces. The vertical loading applied by a hydraulic jack is transferred through the rigid reaction frame and adds on a rigid load plate that is placed on top of the sub-ballast in the upper shear box. The normal load is constant during the tests.

* Corresponding author - Tel:+86-15901173048; fax:+86-10-51683764; E-mail: gjjing@bjtu.edu.cn



FIGURE 1 Shear test box

2.2 BALLAST MATERIAL

Clean sub-ballast samples were prepared in the lower shear box by layers under the conditions similar to the field according to sub-ballast specifications. The sub-ballast material is granite, well compacted before the shearing for each test sample with the same procedure. In the numerical simulation, all the particles are irregularly clumped particles generated according to the sample of tests. The diameters range from 0.075 mm to 45 mm uniformly. The particles generated in the DEM compared with the tests were illustrated by Figure 2.

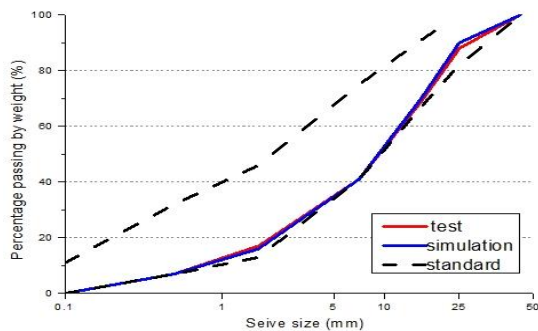


FIGURE 2 Ballast size distribution

2.3 DEM MODEL

Particle shape is one of the most important factors, which influence sub-ballast characteristics in both DEM simulations and experiments, and it is well acknowledged that the granular material (ballast and sub-ballast) shape and size is of great importance for particle interlock ability and rotation characteristics [6-8]. Clumps are efficient and precise compared with simple ball or complex polygons, and latest investigation illustrated that realistic shape to sub-ballast or ballast particles clumps were developed recently [7-10]. For this purpose, overlapping balls are used to form clumps using a simple procedure which gives control over the shape and size of the clumps, small spheres were generated and overlapped assemble according to configuration rules of irregular

particles, by DEM inner language or outer assistance of software. Figure 3 presents several typical ballast particles formed by clump of balls. The box size is more than 20 times of the average particle size, so the size effects could be omitted in the simulation and tests.

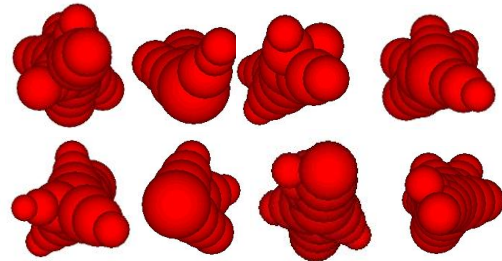


FIGURE 3 Ballast clump particles

The DEM samples were generated whose size and particle size distribution is the same as those of the tests. Once the particles were created, the gravity of the clumps was set as -9.8 m/s^2 . The clumps fell to the bottom of shearing box. An equilibrium state was achieved by cycling the particle assembly. A DEM sample dimension of $30 \times 30 \times 20 \text{ cm}$ was selected and treated as a representative elementary volume (REV), which consisted of 20.2 particles, and the tests' sub-ballast particle number was 57022. When the sample was generated, balanced to the equilibrium state, Then an initial compaction procedure was followed which led the sub-ballast particle assembly to the desired state before loading. All the sub-ballast sample was prepared with the same method for the tests and simulation.

Figure 4 illustrates DEM model of sub-ballast direct shear test numerical model similar to the test conditions. Direct shear condition was obtained by moving the lower boundary walls with a very slow loading rate while maintaining a constant lateral stress, which was achieved by moving the lateral walls inwards or outwards as needed. The constant volume simulation was to guarantee no volume change during the course of shearing. Under the normal force of P , the lower level of the wall was moved with a speed of 0.5 mm/s till a total lateral displacement of 3.5 cm was reached in the DEM simulations. The sub-ballast contact force, number and the shearing force were recorded. The Mohr-Coulomb contact law was used in the simulations based on the ballast DEM simulations [8-10]. The contact parameters are listed on Table 1. The qusai-static and default damping value was used by manual.

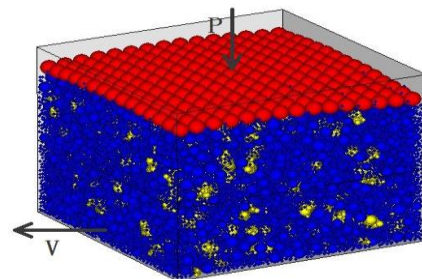


FIGURE 4 DEM model of sub-ballast direct shear test

TABLE 1 Parameters used in numerical simulations

Parameters	Ballast	Wall
Tangential stiffness of particle (N/m)	5×10^8	1×10^9
Normal stiffness of particle (N/m)	5×10^8	1×10^9
Mass density (kg/m ³)		2600
Friction coefficient		0.5
Damping coefficient		0.7

3 Results and Analysis

3.1 SHEAR STRESS AND DISPLACEMENT

Figure 5 shows the recorded shear stresses against the shear displacements for both the laboratory and DEM simulation results under the normal stress of 50 kPa, 100 kPa, 150 kPa, and 200 kPa. Figure 5 indicates the maximum shear strength of 192.6 kPa that occurred at a horizontal displacement of 21.6 mm. The corresponding vertical stress at this deformation was 108 kPa. Thus, a maximum friction coefficient is 0.51. The results above indicate the DEM model predictions are in good agreement with the laboratory results at all normal stress levels for the clean ballast sample.

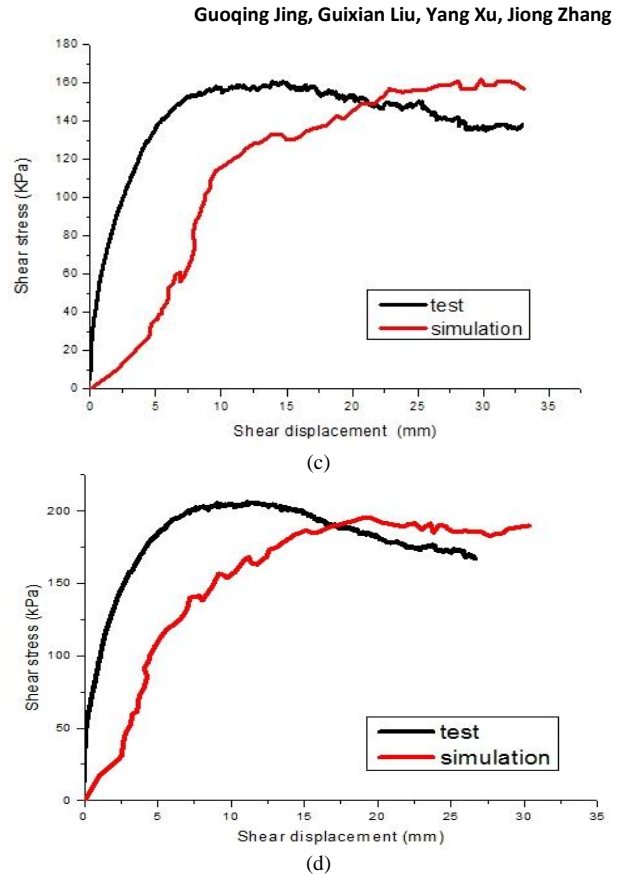
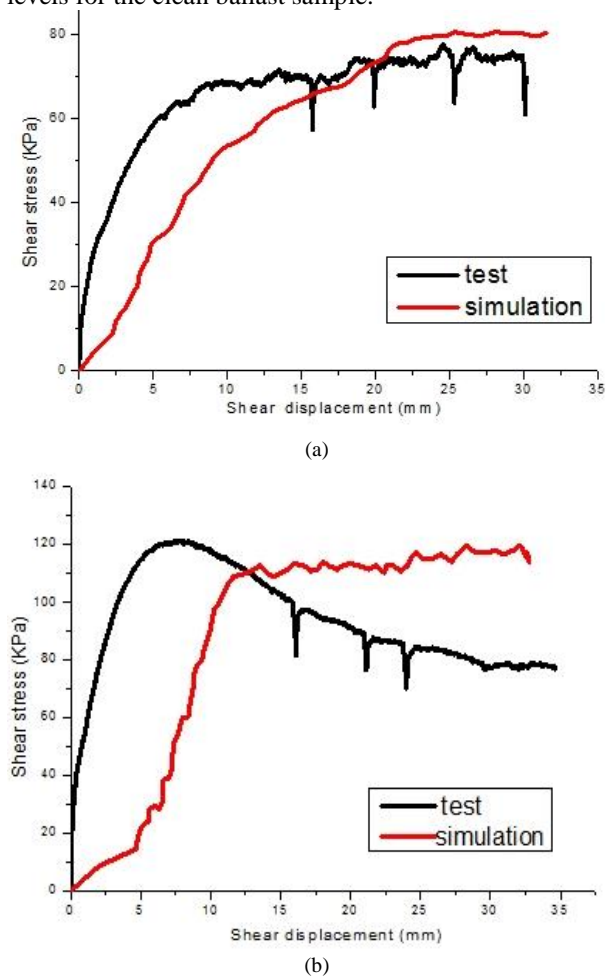


FIGURE 5 Shear force vs. displacement under different normal stress: (a) 50 kPa (b) 100 kPa and (c) 150 kPa (d) 200 kPa

3.2 PARTIAL DISPLACEMENTS

Figure 6 shows the ballast particle displacement vectors for the shear test at the displacement of 2 cm with 30 cm width of shearing box (a section of the box). The DEM results indicate that during the shearing process, the rear of the particles moved downwards with the front particles moving upwards. The movements were interrelated with the ballast contact force chain alteration discussed below. The micro-response of ballast particles under shearing displacements were presented, and could be used to further analyse the ballast-sleeper interaction as well as ballast particle response under sleeper dynamic loading.

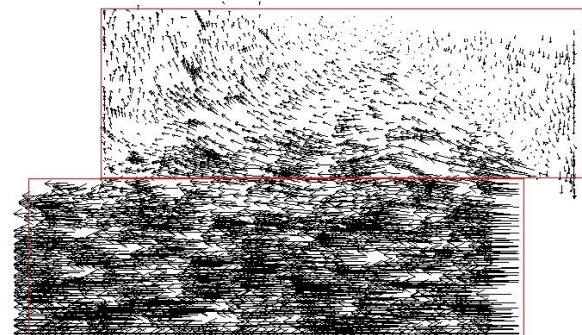
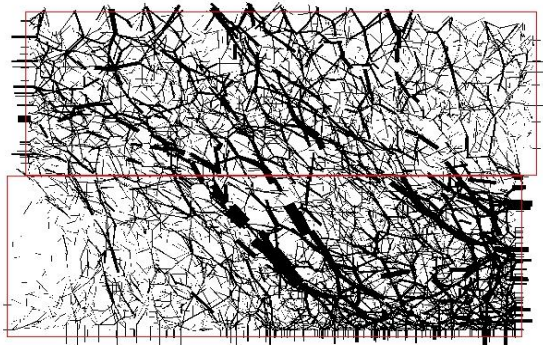


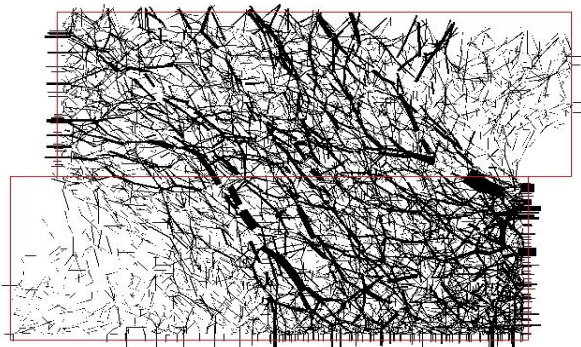
FIGURE 6 Particle displacements (lower box displacement of 2cm)

3.3 CONTACT FORCE CHAIN

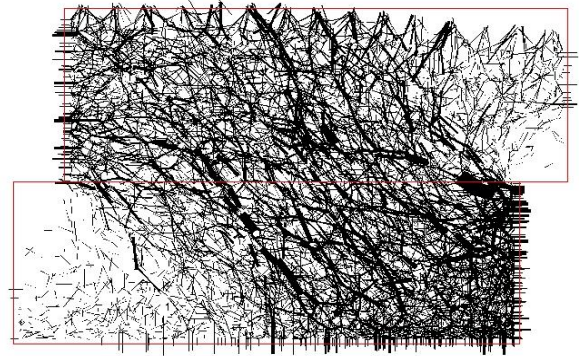
It is of importance to analyse the granular sample direct shear characteristics by investigating the fabric orientation and alteration of sub-ballast particle interlocking and contact. The contact forces were shown as lines with thickness proportional to the magnitude of the contact force. Figure 7 illustrates that with lateral increase of displacements with contact force chain evolutions. At the initial stage with zero wall displacements, the contact force was distributed uniformly throughout the whole assembly, and vertically transmitted, with the process of wall lateral displacements, the contact force chain was altered simultaneously, changing from vertically distributed into horizontally distributed, and the shear bond was observed, and the contact force and contact number were increased up to 3.5 cm displacement [10].



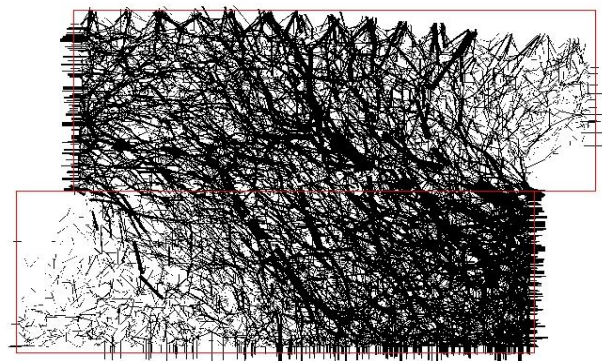
(a) Maximum contacts force: 401.8N; Average contacts force:12.1N; No. of contacts: 391912



(b) Maximum contacts force: 682.7N; Average contacts force: 12.8N; No. of contacts: 391029



(c) Maximum contacts force: 737.2N; Average contact force: 12.7N; No. of contacts: 390969



(d) Maximum contact force: 609.6N; Average contact force: 12.8N; No. of contacts: 390700

FIGURE 7 Contact force for normal stress of 200kPa (a) 1cm; (b) 2cm; (c) 3cm; (d) 3.5cm

The microstructure changes can be represented through the coordination number (CN) and contact unit normal during the loading process. CN is a parameter to quantify and characterize the average contacts of each particle among the whole assemblage. The evolutions of CN during the shear are illustrated in Figure 8. It shows that a big decrease with the CN value occurred immediately after the shearing began. However, bigger CN values were produced under larger stress, indicating that ballast particles during the shear process were compacted. It also shows that CN values of various tests tended to become a constant as shown in Figure 8.

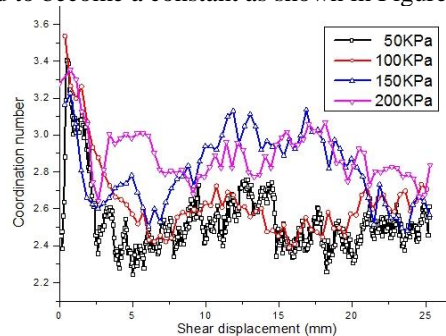


FIGURE 8 CN vs. displacement

3.4 NORMAL STRESS EFFECTS

Figure 9 presents the normal stress effects during the shearing process of sub-ballast material. The results show that the ballast shear stress increased with normal stress, especially before 1 cm of lateral displacements. The average and maximum contact force of shearing ballast, as well as contact number, coordination number under different normal stress were listed in Table 2 with upper box displacement of 1 cm. Table 2 indicates that the average contact force, maximum contact force and contact number increase accordingly as normal stress increase.

TABLE 2 Contacts variation under different normal stress (1cm displacement)

Normal stress (kpa)	Average contact force (N)	Maximum contact force (N)	Contact number	CN
50	7.3	353.1	392763	2.52
100	8.5	448.3	390153	2.61
150	11.9	409.7	387997	2.90
200	12.1	401.8	391912	2.95

Both for laboratory tests and DEM simulations, the forces T (shear force) and N (normal force) indicated as the Figure 9 were used to measure the stress ratio assuming that the $T/N = \tau/\sigma$. In the DEM of direct shear tests, the force ratio was determined from the wall reaction forces shown by Figure 9 by the equation of (1). The results were listed by Table 3, indicated that higher intermediate normal stress of the ballast particles affect the contact density and distribution of contacts that provide lateral support to the force chains, indicating that higher lateral resistance was obtained with higher shear stress. But the stress ratio decreased as the normal stress increased. It indicated that the interaction of shear stress and normal stress was not linear.

$$\frac{T}{N} = \frac{Fn_1 + Fn_2 - Fn_3}{Fn_2 + Ft_1 - Ft_3} \tag{1}$$

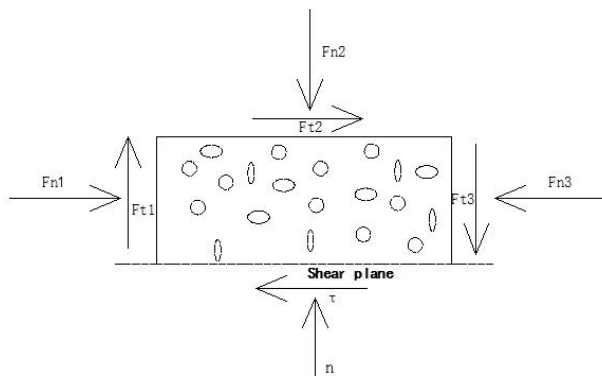


FIGURE 9 Force of the model

TABLE 3 Normal stress effects for the ratio value

Normal stress (kpa)	Maximum shear stress (kpa)	Displacement (mm)	Ratio value
50	78	26	1.56
100	109	27	1.09
150	157	28	1.04
200	197	20	0.985

The sub-ballast samples were sheared horizontally in the shear box under different normal pressures of 50, 100, 150 and 200 kPa, where the relationships between the normal stress and shear stress could be established. The maximum shear stress at failure under each applied normal pressure was recorded from each test, and indicated that the peak shear stress increased with the normal stress increasing.

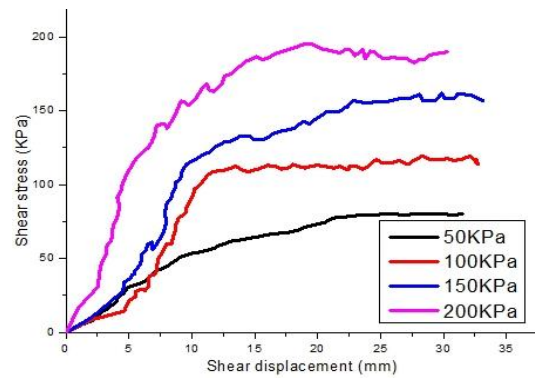


FIGURE 10 Normal stress effects of tests

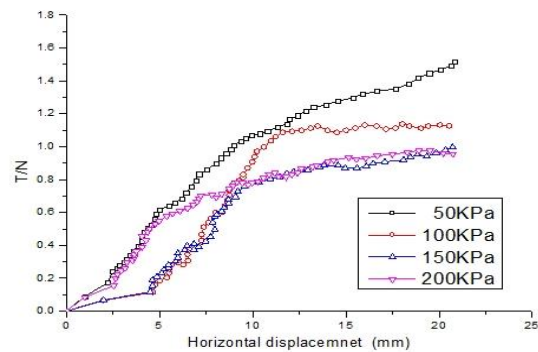


FIGURE 11 T/N vs. horizontal displacement

4 Conclusions and prospective

This paper has studied sub-ballast material response under direct shear tests. The sub-ballast granular is similar ballast, which is fine grained and smaller than ballast. The DEM simulations of biaxial shear test were developed to interpret the evolution of sub-ballast characteristics. The analysis of micro-mechanical characteristics interrelated with the DEM was studied by considering the shape of the particles, size and PSD. The shape property of sub-ballast was constructed based on irregular clumped spheres. The micro-analysis of the laboratory tests under different normal stress levels was

conducted. A satisfactory agreement between discrete analysis and test results was achieved. Some conclusions are given as follows:

DEM successfully simulates the typical granular material behaviour in the direct shear compression tests, as observed from the experimental tests.

The peak angle of shearing resistance increases with the normal stress. A non-linear relationship between normal stress and response characteristics was observed. The shearing force value of sub-ballast was determined under different normal stress level, which could be referred for future track design.

Acknowledgements

This study has been carried out under financial support from NSFC (51108026, 51308324) and Fundamental Research Funds for Central Universities (2014JBM095).

References

[1] Selig E T, Waters J M 1994 *Track technology and substructure management* Thomas Telford, London

[2] Suiker A S J 2005 Static and cyclic triaxial testing of ballast and subballast *Journal of Geotechnical and Geoenvironmental Engineering* 131 771-82
 [3] Ismail and Raymond G P 2003 Monotonic load tests on layered granular aggregates, *Joint Proceedings 55th. Canadian Geotechnical Conference and 2003 North American Geosynthetics Society Conference, Winnipeg, Manitoba* 34 1-8
 [4] Wheat P, Smith A 2008 Assessing the marginal infrastructure maintenance wear and tear costs of Britains railway network *J. Transp. Econ. Policy* 42 189-224
 [5] Lings M L, Dietz M S 2004 An improved direct shear apparatus for sand *Geotechnique* 54(4) 245-56
 [6] Cundall P A, Strack O D L 1979 A discrete numerical model for granular assemblies *Geotechnique* 29 47-65
 [7] Kozicki J, Tejchman J, Mrz Z 2012 Effect of grain roughness on strength, volume changes, elastic and dissipated energies during quasi-static homogeneous triaxial compression using DEM. *Granular Matter* 14(4) 457-68
 [8] Lu M, McDowell G R 2010 Discrete Element Modelling of Railway Ballast under Monotonic and Cyclic Triaxial Loading *Geotechnique* 60(6) 459-67
 [9] Lackenby J, Indraratna B, McDowell G, Christie D 2007 Effect of confining pressure on ballast degradation and deformation under cyclic triaxial loading *Geotechnique* 5 527-36
 [10] Indraratna B, Ngo N, Trung 2011 Behaviour of geogrid-reinforced ballast under various levels of fouling *Geotextiles and Geomembranes* 29(3) 313-22

Authors	
	<p>Guoqing, Jing, born in May, 1979, Beijing, P.R. China</p> <p>Current position, grades: Associate Professor of Beijing Jiaotong University, China. University studies: He received his M.Eng. from Southwest Jiaotong University in China. He received his Ph.D. from INSA de Rennes in France. Scientific interest: His research interest fields include railway ballast, Discrete Element Modelling (DEM). Publications: more than 20 papers published in various journals and one book. Experience: He has teaching experience of 6 years, has completed eight scientific research projects.</p>
	<p>Guixian Liu, born in May, 1988, Beijing, P.R. China</p> <p>Current position, grades: Graduate student of Beijing Jiaotong University, China. University studies: He received his M.Eng. from Shijiazhuang Tiedao University in China. Scientific interest: His research interest fields include railway ballast, Discrete Element Modelling (DEM). Publications: 2 papers published in various journals and conference.</p>
	<p>Yang Xu, born in April, 1987, Beijing, P.R. China</p> <p>Current position, grades: Ph.D. of Beijing Jiaotong University, China. University studies: He received his M.Eng. from Anhui Jianzhu University in China. Scientific interest: His research interest fields include track engineering. Publications: more than 5 papers published in various journals.</p>
	<p>Jiong Zhan, born in August, 1980, Jinan, P.R. China</p> <p>Current position, grades: Associate Professor of Shangdong University, China. University studies: He received his M.Eng. in Environmental Engineering, Ocean University of China. He received his Ph.D. in Civil Engineering, Institute National des Sciences Appliquées (INSA) de Rennes, France. Scientific interest: His research interest fields include, granular material and Computational Fluid Dynamics (CFD). Publications: more than 12 papers published in various journals. Experience: Teaching the Energy and Environment Experiments since 2008.</p>

Numerical analysis and improvement of torsional vibration of shaft systems for engine with cylinder deactivation

Liu Xiaoyong^{*}, Su Tiexiong, Zhang Yi

School of Mechanical and Power Engineering North University of China, Taiyuan City, Shanxi Province, China, 030051

Received 1 January 2014, www.tsi.lv

Abstract

Cylinder deactivation is one of effective ways to improve fuel economy of engine, but will lead to changes in torsional vibration characteristics of shaft systems for engine. A lumped parameter model of torsional vibration of shaft systems for engine with cylinder deactivation was established, the numerical computing method was determined, harmonic analysis was engaged for the excitation torque of single cylinder. Based on these studies, torsional vibration of a V8-engine was analysed, the natural frequency results was verified by comparing with that of utilizing software AMESIM. The forced vibration results show that the torsional angle displacement of crankshaft under cylinder deactivation increases obviously, which mainly consists of the 2rd order rolling vibrations, but torsional stress decreases little. In order to control the rolling vibration, the measure of increasing the rotational inertia of the flywheel was adopted. The results after the adjustment show that the vibration of crankshaft was under control. In a word, the method is feasible and referred.

Keywords: Cylinder Deactivation, Torsional Vibration, Natural Frequency, Forced Vibration, Rolling Vibration, Excitation Torque

1 Introduction

In recent years, cylinder deactivation is widely used to save fuel on some of the advanced internal combustion engine [1]. Cylinder deactivation, when engine working in partial loads, is that the fuel for several cylinders is cut off by some mechanism so that the remainder cylinder for normal combustion can run efficiently in the high load region [2]. Cylinder deactivation would not only reduces pumping losses and mechanical loss, but also improves fuel economy of engine, which the reason is that a low concentrated gas mixture can stabilize combustion due to increasing the mixture gas and reducing residual exhaust gas in the remainder cylinder. When vehicle accelerating fast and climbing, all cylinders will start work to enhance engine power output. Currently, the measure for the deactivating cylinder is now widely used by cutting off fuel and closing valve [3].

The fuel will be saved as much as 20% by employing the engine with cylinder deactivation. Nevertheless, under cylinder deactivation, the firing interval angle of remainder cylinder and the unevenness of running will increase. There will be changes in torsional vibration characteristics of rotating shaft systems for engine. The rotating crankshaft will produce torsional vibration due to the effects of periodic excitation torque, which torsional vibration is a common vibration form of rotating machinery. When excitation torque frequency and natural frequency of crankshaft are equal, the resonance in shafting be evoked. The resonance would

have invoked a great stress of torsional vibration that can cause damage to the crankshaft [4]. The situation must be under control. Consequently, numerical analysis of torsional vibration for engine is very necessary.

Torsional vibration characteristics of shaft systems for engine with several cylinders deactivation was studied in this study. For the aim of providing a theoretical reference for the applications of cylinder deactivation in engines by taking a kind of V8-engine for example, which the engine would run in 8 cylinder, or 4 cylinder, The impact of torsional vibration for engine with and without cylinder deactivation were comparatively analysed. Improvement measures were proposed.

2 Modelling and numerical computing methods

2.1 LUMPED PARAMETER PHYSICAL MODELS

The actual structure of rotating shafts of engine is very complex elastic motion bodies, as shown in Figure 1. According to the principle of energy conservation, they have been addressed as a lumped parameter physical model and simplified as discretized systems with several lumped mass discs in which consists connected massless elastic shaft elements [5]. Simplified equivalent model for torsional vibration of shaft systems was shown as Figure 2, which the lumped mass number represents different part in shaft systems. M_m represents excitation torque of crank.

^{*} Corresponding author - Tel: +86-351-392-0980; fax: +86-351-392-0980; E-mail: liuxiaoyong0354@nuc.edu.cn

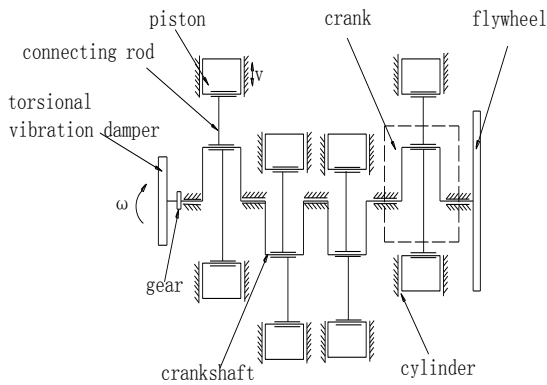


FIGURE 1 Sketch of shafts systems

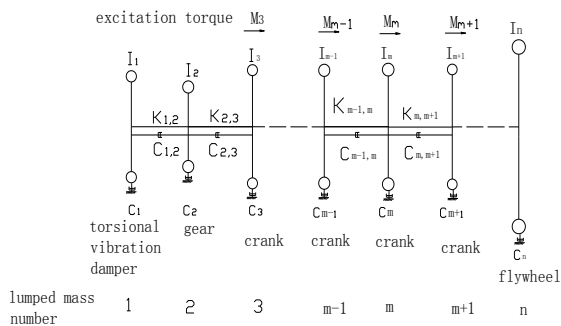


FIGURE 2 Lumped parameter model

2.2 MATHEMATICAL MODEL

The multi-mass lumped parameter model is a multi-degree of freedom torsional vibration system. The kinetic equations of the *m*-th mass is expressed according to d'Alembert principle as follows.

$$I_m \ddot{\varphi}_m(t) + C_m \dot{\varphi}_m(t) + C_{m-1,m} (\dot{\varphi}_m(t) - \dot{\varphi}_{m-1}(t)) + C_{m,m+1} (\dot{\varphi}_m(t) - \dot{\varphi}_{m+1}(t)) + K_{m-1,m} (\varphi_m(t) - \varphi_{m-1}(t)) + K_{m,m+1} (\varphi_m(t) - \varphi_{m+1}(t)) = M_m \quad (1)$$

where $\varphi_m(t)$, $\dot{\varphi}_m(t)$, $\ddot{\varphi}_m(t)$, I_m , C_m , M_m , ψ_v are respectively torsional angle displacement, angular velocity, angular acceleration, rotational inertia, viscous damping factor, excitation torque, excitation torque amplitude, initial phase angle for the *m*-th lumped mass; $K_{m-1,m}$ and $C_{m-1,m}$ are respectively torsional stiffness and damping coefficient of shaft element *m-1*, *m*; *t* is the time.

Each lumped mass has the same mathematical expression, and *n* lumped masses can compose *n* differential equations. It can be written in matrix form.

$$I \ddot{\varphi}(t) + C \dot{\varphi}(t) + K \varphi(t) = M(t) \quad (2)$$

where *I* is the inertia matrix; *C* is the damping matrix; *K* is the stiffness matrix; *M*(*t*) is the excitation torque vector; $\varphi(t)$ is the torsional angle displacement vector.

2.3 EXCITATION TORQUE MODEL FOR ENGINE

Excitation torque with the periodically change is the energy source of torsional vibration resulted from the tangential force imposed on the crankpin of crank.

2.3.1 Excitation torque for normal combustion cylinder

Excitation torque of single cylinder under normal combustion cylinder is

$$M = M_p + M_j + M_g \quad (3)$$

where M_p , M_j and M_g is the excitation torque on crank of single cylinder pressure, reciprocating inertial force and moving bodies gravity, respectively. written by

$$M_p = \frac{\pi D^2}{4} p R \frac{\sin(\alpha + \beta)}{\cos \beta} \quad (4)$$

$$M_j = -m_j \omega^2 R^2 (\cos \alpha + \lambda \cos 2\alpha) \frac{\sin(\alpha + \beta)}{\cos \beta} \quad (5)$$

$$M_g = m_j g R \frac{\sin(\alpha + \beta)}{\cos \beta} \quad (6)$$

In Figure 3, where *p* is the cylinder pressure, which is obtained by the experiment or the simulation. *D* is the cylinder bore, *R* is the crank radius, α is the crank angle, β is the angle between the rod and the centre line of cylinder, m_j is the bodies mass with the reciprocating motion, λ is the crank and rod ratio, ω is the crank angular velocity.

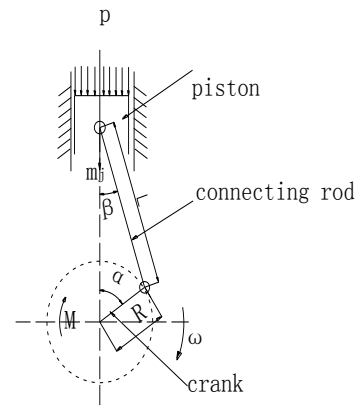


FIGURE 3 Sketch of crank rod system

2.3.2 Excitation torque for deactivating cylinder

Only compression and expansion of initial air took place in deactivating cylinder due to the close of intake and exhaust valve. The excitation torque on the crank imposed by cylinder pressure is very small and could be neglected compared to explosive pressure in normal combustion cylinder. Nevertheless, the excitation torque produced by reciprocating inertial force and moving bodies gravity still remains with piston mechanism of deactivating cylinder reserved. Thus, excitation torque of

single cylinder for deactivating cylinder would be written by

$$M = M_j + M_g . \tag{7}$$

The excitation torque curve of single cylinder in a working cycle of a V8 engine under declared working condition was showed in Figure 4. Figure 4(a) and Figure 4(b) are respectively excitation torque for normal combustion cylinder and deactivating cylinder.

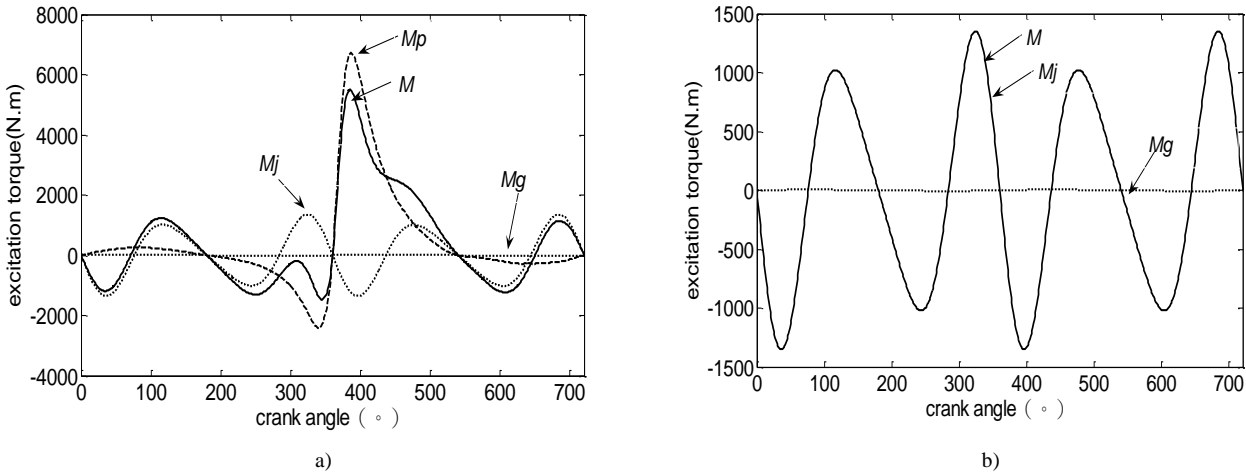


FIGURE 4 Excitation torque of single cylinder in a working cycle: a) Normal combustion cylinder, b) Deactivating cylinder

2.3.3 Harmonic analysis for excitation torque

The periodic excitation torque of single cylinder can be expressed as Fourier series [6], which is equal to the sum of average torque and harmonic torque with different amplitude and different frequency, written by

$$M = a_0 + \sum_v M^v \sin(v\omega t + \psi_v), \tag{8}$$

where a_0 is the average torque, v is the order, ψ_v is the phase angle, M^v is the excitation torque amplitude under the v -th order.

Each cylinder on the crankshaft in multi-cylinder engine works in a regular firing order. The crankshaft would be affected by several excitation torque, which have a certain amount of phase angle difference between each other. There are two cylinders fixed on one single crank on V-engine, so excitation torque of v -th order on one single crank is written by:

$$M^v = M_R^v \sin(v\omega t + v\beta_{1,R}) + M_L^v \sin(v\omega t + v\beta_{1,L}), \tag{9}$$

where M_R^v and M_L^v is respectively the excitation torque of the v -th order of left cylinder and right cylinder on this single crank. where $\beta_{1,L}$ and $\beta_{1,R}$ is respectively the firing interval angle of left and right cylinders compared to the first cylinder.

2.4 NUMERICAL COMPUTING METHODS

2.4.1 Numerical computing method of free vibration

Free vibration is a vibration without excitation torque. The computation aims to solve the natural frequency and mode shape. Assuming that each lump mass functions a harmonic vibration, then the torsional angle displacement vector can be represented as:

$$\varphi(t) = Ae^{\omega t}, \tag{10}$$

$$A = [A_1, A_2, \dots, A_n]^T, \tag{11}$$

where A is mode shape vector, A_1, A_2, A_n are the amplitude of torsional angle displacement for the first, second and n -th lumped mass.

Substitute equation (10) into equation (2) and let the excitation torque and damping coefficient in the equation (2) equals to zero, which damping is neglected thanks to little influence to free vibration. The problem can be changed into solving the problem of eigenvalue of matrix [7].

$$(\mathbf{K} - \omega^2 \mathbf{I})Ae^{\omega t} = 0. \tag{12}$$

By solving the eigenvalue and eigenvector of matrix K/I , The natural frequency and mode shape can be obtained.

2.4.2 Numerical computing method of forced vibration

The computation of forced vibration aims to solve the torsional vibration response of shaft systems to excitation torque. The solution of the m -th lumped mass under the influence of harmonic excitation torque of the v -th order is written by

$$\varphi_m^v = A_m^v e^{iv\omega t + \varepsilon_m^r}, \tag{13}$$

where A_K^v and ε_K^r is respectively the amplitude and phase angle of torsional angle displacement for lumped mass m of v -th order. Substitute the solution for each lumped mass into equation (2) to get n united complex number equations. Decompose each complex number into the real part and the imaginary part and get the $2n$ united real number linear equations. A_m^v and ε_K^r could be obtained after solving these equations.

According to the principle of linear superposition, the torsional angle displacement synthesis amplitude for the m -th lumped mass is written by

$$A_m = \sum_v A_m^v e^{iv\omega t + \varepsilon_m^r} = \sum_v A_m^v \sin(v\omega t + \varepsilon_m^v). \tag{14}$$

The torsional stress synthesis amplitude for shaft element $m, m+1$ is got by

$$A\tau_{m,m+1} = K_{m,m+1} \sum_v (A_m^v - A_{m+1}^v) / W_{m,m+1}, \tag{15}$$

where $W_{m,m+1}$ is the section modulus in torsion for shaft element $m, m+1$.

3 Numerical analysis and simulations

Torsional vibration of shaft systems was analysed with a V8 four-stroke engine under normal combustion of all cylinders and 4-cylinder deactivation. The V firing angle of the engine is 90^0 . The firing order is R1-L4-L2-R2-L3-R3-R4-L1. Considered the uniformity of the output torque of the engine, deactivating cylinders are R1, L2, L3 and R4 under the 4-cylinder deactivation. The lumped parameter model of shaft systems of the engine is established as shown in Table 1.

TABLE 1 Lumped parameter model of shaft systems

Lumped mass number	Rotational inertia (kg.m2)	Shaft number	torsional stiffness (N.m/rad)
1	0.479	1,2	0.22e6
2	0.185	2,3	5.404e6
3	0.246	3,4	3.192e6
4	0.246	4,5	3.192e6
5	0.246	5,6	3.192e6
6	0.246	6,7	6.606e6
7	5.786		

3.1 COMPUTATION OF NATURAL FREQUENCY

The natural frequency of shaft systems are same with normal combustion of all cylinders and under cylinder deactivation since there is no change in the structure of shaft systems. The natural frequency results are shown in Table 2. To verify the accuracy of the results, the simulation software AMESIM, which is widely used to simulate in the mechatronics and hydraulics domain, was used for the calculation of natural frequency of shaft systems for the engine. As shown in Table 2, the calculation error is less than 3%, consequently the results' reasonability and credibility of the calculating program was verified.

TABLE 2 Lumped parameter model of shaft systems

Model No.	Based on this study (r/min)	Based on AMESIM (r/min)	Error (%)
1	5890	5888	0.003
2	13741	13634	0.7
3	34610	34020	1.9
4	54117	53524	1.1
5	67245	66414	1.3
6	74829	74214	0.8

3.2 COMPUTATION OF FORCED VIBRATION

3.2.1 Engine with normal combustion of all cylinders

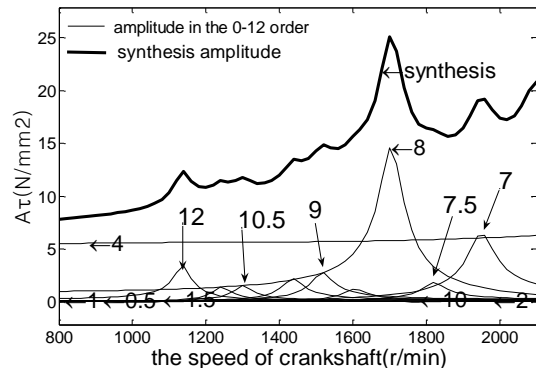
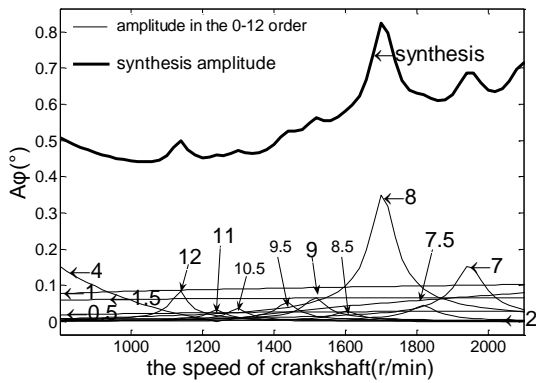
Numerical computation of forced vibration for engine with normal combustion is displayed in Figure 5. From Figure 5(a), the synthesis amplitude of torsional angle displacement on the free end of crankshaft (the second lumped mass) can reach its maximum as much as 1700r/min. The resonance will happen when the frequency of excitation torque of the order 7, 7.5, 8, 8.5, 9, 9.5, 10.5, 11 and 12 equals to the 2rd natural frequency of shaft systems. The peak amplitude of the order 8 is the highest among all orders.

From Figure 5(b), the torsional stress synthesis amplitude on shaft element 6,7 (between crank 4 and flywheel) reaches its maximum value when this crankshaft speeds up to 1700r/min. The peak torsional stress of the order 8 is the highest among all orders.

3.2.2 Engine under cylinder deactivation

From Figure 6(a), under 4-cylinder deactivation, the synthesis amplitude of torsional angle displacement on the free end of crankshaft sharply increases. The peak amplitude 1.1 is 2.4 times as large as that under normal combustion in the corresponding speed. The main cause is there appears rolling vibration of the order 2 in low speed region. Compared with that under normal combustion, the torsional angle displacement of all other orders except for the order 2 and 10 decreases.

From Figure 6(b), even though rolling vibration forces a augmentation to the amplitude of torsional angle displacement on free-end, the torsional stress amplitude in the order 2 is not very big because the amplitude difference of torsional angle displacement between two lumped mass at the two ends of shaft element 6,7 has not increased a lot. Compared with that under normal combustion, the torsional stress amplitude of the synthesis and all other orders except for the order 2 and 10 will decrease because of going up of even firing interval angle under 4-cylinder deactivation and reduce the number of normal combustion cylinders.

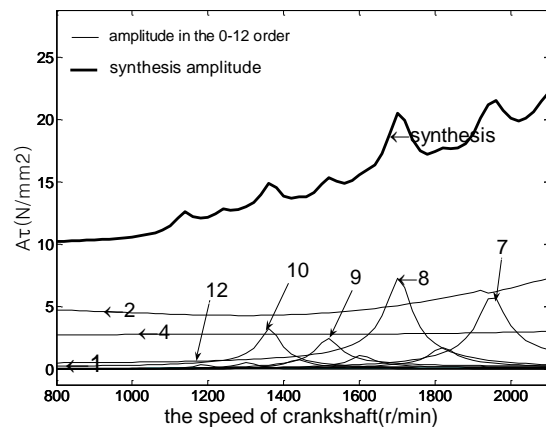
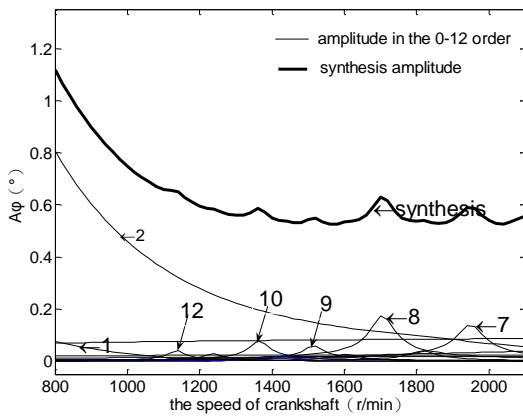


a)

b)

FIGURE 5 Torsional vibration of engine with normal combustion:

a) Torsional angle displacement amplitude on the free end of crankshaft, b) Torsional stress amplitude on shaft element 6,7



a)

b)

FIGURE 6 Torsional vibration of engine under 4-cylinder deactivation:

a) Torsional angle displacement amplitude on the free end of crankshaft, b) Torsional stress amplitude on shaft element 6,7

4 Improvement measures

Rolling vibration is that torsional angle displacement amplitude of all lumped masses swings with the same value and direction. According to equation (15), Rolling vibration produce no torsional stress and is harmless to strength of shafts, but excessive amplitude of rolling vibration will worsen the motion characteristics and dynamic load for the valve train driven by the crankshaft and ultimately cause damage to it. The situation must be

under control. From Figure 6(a), the amplitude of the 2 order decreases with increase of excitation frequency (the product of the order and the speed of crankshaft). The rolling vibration amplitude of the order 2 is smaller when excitation frequency is located closer to the 1st natural frequency (5900r/min). However, rolling vibration amplitude is not obvious in the order 0.5, 1, 1.5, the reason is that excitation torque vector sum of all the cylinders in the above orders is relatively small. It indicates that rolling vibration amplitude have

relationship with natural frequency of shaft systems and vector sum of all the cylinders. Above all, rolling vibration amplitude can be controlled by reducing natural frequency and excitation torque vector sum of all the cylinders.

In point of the structural characteristics of the V8-engine, the rolling vibration amplitude can be controlled by reducing the natural frequency through increasing the rotational inertia of the flywheel. The rotational inertia can be changed by adjusting its thickness. When the rotational inertia of the flywheel increase to 10.786kgm^2 ,

numerical computation of forced vibration for engine under 4-cylinder deactivation is shown in Figure 7.

Compared the result before the adjustment, from Figure 7(a), the torsional angle displacement amplitude of the synthesis and the 2-th order will decrease at the speed of 800r/min. Torsional stress amplitude changes little from Figure 7(b). The amplitude of the 2 order is lower than allowable amplitude 5° [8]. It indicates that the method of reducing the rolling vibration amplitude by adjusting the inertia of the flywheel is feasible.

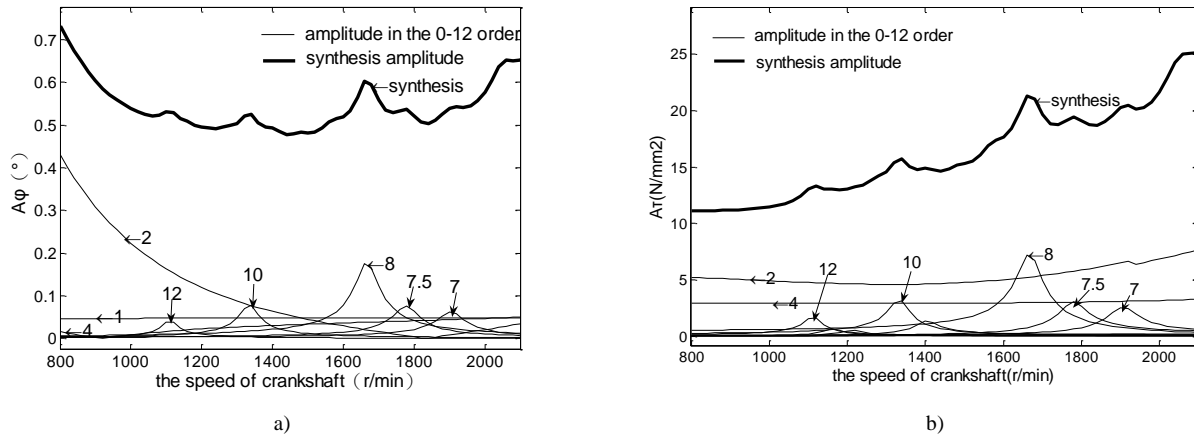


FIGURE 7 Torsional vibration of engine under 4-cylinder deactivation after the adjustment:

a) Torsional angle displacement amplitude on the free end of crankshaft, b) Torsional stress amplitude on shaft element 6,7

5 Conclusion

Through the numerical computation for torsional vibration of shaft systems for engine with deactivating cylinder, conclusions drawn as below: The natural frequency characteristics of shaft systems do not change under cylinder deactivation but only change of that of forced vibration. For a V8-engine under the normal combustion of all cylinders, the torsional angle displacement amplitude of order 8 are the biggest of all orders, there is no rolling vibration within the working speed. Under 4-cylinder deactivation, the torsional angle displacement amplitude had an obvious augmentation, which mainly consists of order 2 rolling vibrations, but the torsional stress synthesis amplitude decreased a little. In order to reduce the rolling vibration amplitude, the measure of reducing the natural frequency by increasing the rotational inertia of the flywheel was adopted. The results after the adjustment showed that the rolling vibration decreases and torsional stress amplitude changes little. This study contribute to comprehend the torsional vibration characteristics and provide

improvement measures to control vibration for engine with cylinder deactivation.

References

- [1] Kai W, Tingfang Z, Zhijun W, Jun D, Tao W 2010 Optimizing on the simulation of the gasoline engine with cylinder deactivation *Small Internal Combustion Engine Motorcycle* **39**(1) 84-7
- [2] Osman A K, Hikmet A, Alper T C 2005 Methods to improve efficiency of four stroke, spark ignition engines at part load *Energy Conversion and Management* **46**(20) 3202-20
- [3] Dengpan Z, Yinnan Y, Yong C 2007 Technology of cylinder deactivation of automotive gasoline engine *Small Internal Combustion Engine Motorcycle* **36**(6) 89-93
- [4] Li H, Stone B J 1999 Time domain modelling of a reciprocating engine *Mechanical Systems and Signal Processing* **13**(1) 169-78
- [5] Qing H, Dongmei D 2010 Modelling and calculation analysis of torsional vibration for turbine generator shafts *Journal of Information & Computational Science* **10**(7) 2174-82
- [6] Fuchun Y, Zhenxing S, Jianfeng M 2013 Nonlinear dynamics and load sharing of double-mesh helical gear train *Journal of Engineering Science and Technology* **6**(2) 29-34
- [7] Jianhua X, Ridong L, Weizhen Z 2007 Structural dynamic modification of crankshaft torsional vibration system of IC engine base on sensitivity analysis *Chinese Internal Combustion Engine Engineering* **28**(6) 66-9 (in Chinese)
- [8] Zhaohan W 1990 *Internal combustion engine design* Press: Beijing, 175-95 (in Chinese)

Authors



Xiaoyong Liu, born in February, 1980, Pingyao County, Shanxi Province, P.R. China

Current position, grades: the lecturer of School of Mechanical and Power Engineering, North University of China, China.

University studies: received his master of engineering from North University of China in China.

Scientific interest: Scientific interest: His research interest fields include numerical analysis of torsional vibration for engine, gear tooth modification design.

Publications: more than 4 papers published in various journals.

Experience: He has teaching experience of 10 years, has completed 3 scientific research projects.



Tiexiong S, born in November, 1963, Yuanping County, Shanxi Province, P.R. China

Current position, grades: the Full Professor of College of Mechatronic Engineering, North University of China, China.

University studies: received his doctor of engineering from Beijing Institute of Technology in China. He received his master of engineering from Beijing Institute of Technology in China.

Scientific interest: His research interest fields include modern design theory and manufacture of power engineering.

Publications: more than 120 papers published in various journals.

Experience: He has teaching experience of 28 years, has completed 50 scientific research projects.



Yi Zhang, born in January, 1969, Pingshan County, Hebei Province, P.R. China

Current position, grades: the Associate of Professor of Mechanical and Power Engineering, North University of China, China.

University studies: received his Master of Engineering from North University of China in China.

Scientific interest: His research interest fields include structural strength and reliability of engine, modern design methods of engine

Publications: more than 10 papers published in various journals.

Experience: He has teaching experience of 20 years, has completed 13 scientific research projects.

Simulation design of bursa-wolf coordinate transformation model based on the access

Gong Yusheng*

School of Civil Engineering, University of Science and Technology Liaoning, Anshan, China

Received 1 March 2014, www.tsi.lv

Abstract

With the popularity of GNSS technology, problems of conversion between different spatial rectangular coordinate systems are often encountered, which are difficult to solve. This paper introduces Bursa-Wolf coordinate transformation model in detail, this model is widely used in the conversion between different spatial rectangular coordinate system, In order to improve the efficiency of calculation, This paper writes a simulation program using tables, forms, macros and VBA in Access, also the simulation program has been verified by production instances, and the reliability of its results is ensured.

Keywords: Access, VBA, Spatial rectangular coordinate system, Bursa-Wolf Model, Simulation

1 Introduction

With the development of modern surveying and mapping science and technology, especially the emergence of space satellite navigation and positioning technology, the conventional geodetic methods have been gradually replaced by satellite geodetic methods, the location of ground points can be measured and expressed in a three-dimensional geocentric coordinate system by means of satellite geodesy. Both can be expressed as three-dimensional space rectangular coordinate, and earth latitude and longitude of an ellipsoid corresponding to the earth, Based on the above, space geodetic rectangular coordinate system and its transformation in modern geodetic surveying carries more practical significance [1].

There are certain connections between the space rectangular coordinate system and the geodetic coordinate system in one ellipsoid – They can convert to each other. The representing forms of coordinate of a same point are different in two coordinate systems, this is just two different but mutually equivalent representing forms. But when measured in space technologies including GNSS positioning technology, conversions between different datum is often needed, for example, the conversion between the two different geocentric coordinate systems – WGS84 and ITRF; and the conversion between two different ellipse-centered coordinate systems – Beijing geodetic coordinate system 1954 and the 1980 national geodetic coordinate system [2-6].

2 Bursa-Wolf transformation model

Bursa-Wolf transformation model (which is often called Bursa model in short) is also called 7-Parameter Transformation or 7-Parameter Helmert Transformation, as is shown in Figure 1. There are 7 Parameters in the model, which contains 3 Translation parameters: T_X , T_Y , T_Z and 3 Rotation parameters: ω_X , ω_Y , ω_Z (also known as the 3 Euler angles) and 1 scale parameter m [7-8].

Suppose that there are two space rectangular coordinate system based on different datum $O_A-X_A Y_A Z_A$ and $O_B-X_B Y_B Z_B$, Bursa-Wolf model can be used to transform coordinate of $O_A-X_A Y_A Z_A$ to coordinate of $O_B-X_B Y_B Z_B$ by the following steps [9]:

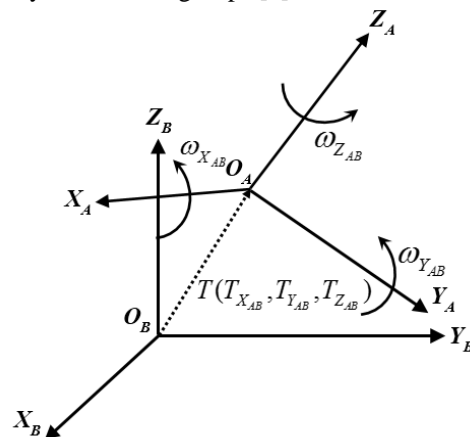


FIGURE 1 Bursa - Wolf seven-parameter transformation

(1) Look at the origin O_A from the side of X_A , O_A is a fixed point in rotate process, the $O_A-X_A Y_A Z_A$ around the X_A axis is counterclockwise rotated $\omega_{X_{AB}}$, so that axis Y_A after rotated is parallel to the $O_B-X_B Y_B$ plane;

* Corresponding author - Tel: +86-412-592-9611; fax: +86-412-592-9610; E-mail: gys_long@qq.com

(2) Look at the origin O_A from the side of Y_A , point O_A is a fixed point in rotate process, the $O_A-X_A Y_A Z_A$ around the Y_A axis is counterclockwise rotated $\omega_{Y_{A,B}}$, so that axis X_A after rotated is parallel to the $O_B-X_B Y_B Z_B$ plane. Obviously, the Z_A axis parallel with the Z_B axis;

(3) Look at the origin O_A from the side of Z_A , point O_A is a fixed point in rotate process, the $O_A-X_A Y_A Z_A$ around the Z_A axis is counterclockwise rotated $\omega_{Z_{A,B}}$, so that axis X_A after rotated is parallel to the X_B . Obviously, three axes of $O_A-X_A Y_A Z_A$ are parallel to three axes of $O_B-X_B Y_B Z_B$;

(4) Zoom the unit of length in the $O_A-X_A Y_A Z_A$ $1+m$ times so that it is the same as the unit of length in the $O_B-X_B Y_B Z_B$;

(5) Respectively move the origin of $O_A-X_A Y_A Z_A$ along X_A axis, X_B axis and X_Z axis $-T_{X_{A,B}}, -T_{Y_{A,B}}, -T_{Z_{A,B}}$ so that it coincides with the origin of $O_B-X_B Y_B Z_B$. Formulas can be used to express the conversion process as follows:

$$\begin{bmatrix} X_B \\ Y_B \\ Z_B \end{bmatrix} = \begin{bmatrix} T_{X_{A,B}} \\ T_{Y_{A,B}} \\ T_{Z_{A,B}} \end{bmatrix} + (1+m_{A,B})R_3(\omega_{Z_{A,B}})R_2(\omega_{Y_{A,B}})R_1(\omega_{X_{A,B}}) \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} \quad (1)$$

In the formula:

$$R_1(\omega_{X_{A,B}}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_{X_{A,B}} & \sin \omega_{X_{A,B}} \\ 0 & -\sin \omega_{X_{A,B}} & \cos \omega_{X_{A,B}} \end{bmatrix}$$

$$R_2(\omega_{Y_{A,B}}) = \begin{bmatrix} \cos \omega_{Y_{A,B}} & 0 & -\sin \omega_{Y_{A,B}} \\ 0 & 1 & 0 \\ \sin \omega_{Y_{A,B}} & 0 & \cos \omega_{Y_{A,B}} \end{bmatrix}$$

$$R_3(\omega_{Z_{A,B}}) = \begin{bmatrix} \cos \omega_{Z_{A,B}} & \sin \omega_{Z_{A,B}} & 0 \\ -\sin \omega_{Z_{A,B}} & \cos \omega_{Z_{A,B}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Take the normal circumstances into consideration, Three Euler angles of $\omega_X, \omega_Y, \omega_Z$ of rotation between the two different datum are all very small, so Bursa-Wolf model can eventually be simplified and shown as following [10-12]:

$$\begin{bmatrix} X_B \\ Y_B \\ Z_B \end{bmatrix} = \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & -Z_A & Y_A & X_A \\ 0 & 1 & 0 & Z_A & 0 & -X_A & Y_A \\ 0 & 0 & 1 & -Y_A & X_A & 0 & Z_A \end{bmatrix} \begin{bmatrix} T_X \\ T_Y \\ T_Z \\ \omega_Z \\ \omega_Y \\ \omega_X \\ m \end{bmatrix} \quad (2)$$

3 Solving for seven parameters

While do Spatial coordinate transformation, if the number of control points n is greater than 3, seven parameters of spatial coordinates model can be solved by using indirect adjustment method, as the following method:

3.1 COLUMN ERROR EQUATION

Coordinates of B coordinate system can be considered as observations, suppose coordinates of B coordinate system are error-free, and then error equation can be columned

$$\begin{bmatrix} v_{x_1} \\ v_{y_1} \\ v_{z_1} \\ \vdots \\ v_{x_n} \\ v_{y_n} \\ v_{z_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -Z_1 & Y_1 & X_1 \\ 0 & 1 & 0 & Z_1 & 0 & -X_1 & Y_1 \\ 0 & 0 & 1 & -Y_1 & X_1 & 0 & Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & -Z_n & Y_n & X_n \\ 0 & 1 & 0 & Z_n & 0 & -X_n & Y_n \\ 0 & 0 & 1 & -Y_n & X_n & 0 & Z_n \end{bmatrix} \begin{bmatrix} T_X \\ T_Y \\ T_Z \\ \omega_Z \\ \omega_Y \\ \omega_X \\ m \end{bmatrix} - \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ \vdots \\ X_n \\ Y_n \\ Z_n \end{bmatrix}_B - \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ \vdots \\ X_n \\ Y_n \\ Z_n \end{bmatrix}_A \quad (3)$$

Written in matrix form $V = BX - L$

Coordinates of each point can be regarded as independent observations of same accuracy, so $P=I$.

3.2 SOLVING FOR PARAMETERS

Put the known coordinate's values of each point into the above error equation, and then solving parameter values according to the following formula:

$$\hat{X} = (B^T B)^{-1} (B L)$$

4 Simulation program design

4.1 CREATE TABLES

3 tables have been designed in the program, structures of each table are shown in table 1-3. Common points data of calculating seven parameters can be saved to the common point table of calculating seven-parameter space rectangular coordinates; Via using the Table of space rectangular coordinate before the seven-parameter conversion, data calls can be achieved, and data storage can also be achieved via using converted seven parameters table [13-14].

TABLE 1 The calculation of common point Tab structure from spatial rectangular coordinate through seven parameters

Field name	Field type	Field length	Decimal digits
Point name	text	50	-
A ellipsoid X	figure	double precision	4
A ellipsoid Y	figure	double precision	4
A ellipsoid Z	figure	double precision	4
B ellipsoid X	figure	double precision	4
B ellipsoid Y	figure	double precision	4
B ellipsoid Z	figure	double precision	4

TABLE 2 The Tab structure from spatial rectangular coordinate before seven-parameter conversion

Field name	Field type	Field length	Decimal digits
Point name	text	50	-
X	figure	double precision	4
Y	figure	double precision	4
Z	figure	double precision	4

TABLE 3 The Tab structure from spatial rectangular coordinate after seven-parameter conversion

Field name	Field type	Field length	Decimal digits
Point name	text	50	-
X	figure	double precision	4
Y	figure	double precision	4
Z	figure	double precision	4

4.2 CREATE FORM OF CALCULATION

Through the form of seven parameters we can achieve:

(1) input space transformation coordinates of the common point, click the command button of “calculate seven parameters” seven transformation parameters can be calculated, as is shown in Figure 2;

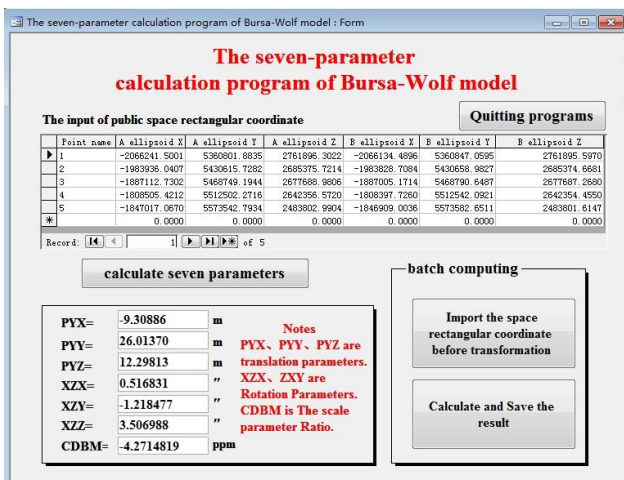


FIGURE 2 The interface of seven-parameter calculation

(2) It can bring out the space rectangular coordinate table before transformation by click the command button of “Import space rectangular coordinate before transformation” and editing it, as is shown in Figure 3, then click the command button of “calculate and save the result” to go on batch conversion, save the calculated data in the converted space transformation coordinate table as well as the Excel table, and eventually saved in the computer - specified directory.

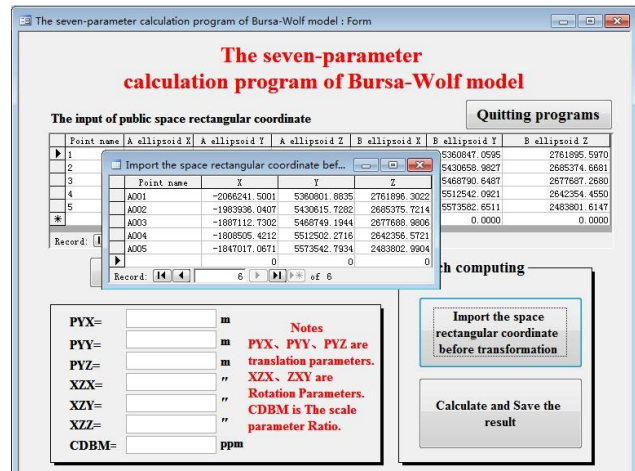


FIGURE 3 The interface of batch calculation

4.3 CREATING MACROS

Using Macros can realize some auxiliary functionalities flexibly in programming, therefore, the Date Input and Export Macro Group is being used in the program design, there into, the export of The Space Rectangular Coordinate Table before The Seven Parameters Transformation can be realized through the use of The Import the Space Rectangular Coordinate Macro before Transformation, the Macro design is shown on Figure 4; Computed data can be saved in The space rectangular coordinate table and also be saved in excel table which eventually being stored in the computer – specified directory, the Macro Design is shown on Figure 5. Lots of programming can be saved just need to use The DoCmd Method to invoke it [15-16].

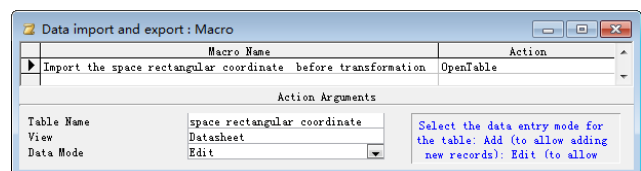


FIGURE 4 Importing the macro of spatial rectangular coordinate before converting

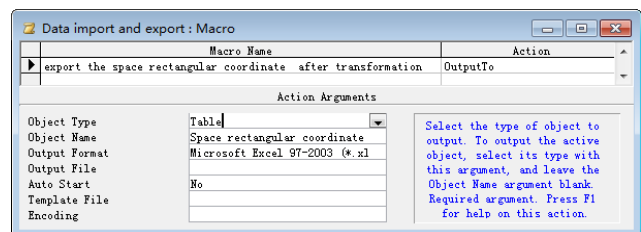


FIGURE 5 Exporting the macro of spatial rectangular coordinate after converting

4.4 CREATING PUBLIC MODULES

In order to make it convenient to realize the Bursa - Wolf Seven Parameters Model Calculation Program, commonly used functions is written into programs and

saved in public modules, that will also be convenient for the invoke while the windows program is running, the main content are as follows [17-20]:

- 'General process of matrix transpose
- Public Sub MatrixTrans(A, c)
- 'Matrix multiplication
- Public Sub MatrixMulty(Qn,Qa,Qb)
- 'List out the general procedures of the Gaussian Pivot Element Reduction Method solving the Liner System of Equations
- Public Sub MajorInColGuass(A,B,CX)
- 'Commonly indirect adjustment
- Public Sub InAdjust(A,P,L,CX)

5 Simulation program verification

In order to verify the reliability of the program, two methods were used in production instances to calculate the parameters of the conversion between the WGS84 and Beijing geodetic coordinate system 1954. One is using the established formula and MATLAB together, the other is using the self-programming which is mentioned above. As is shown on table 5, the results are exactly the same, which prove the accuracy of the self-programming.

TABLE 4 Coordinates in the WGS-84, and BJ-54 coordinate systems

Point	X84	Y84	Z84	X54	Y54	Z54
1	-206241.5001	5360801.8835	27618963022	-20661344896	53608470895	27618955970
2	-19839360407	54306157282	26853757214	-19838287084	54306889827	26853746681
3	-18871127302	54687491944	26776889806	-18870051714	54687906487	26776872680
4	-18085054212	55125022716	26423565720	-18083977260	55125420921	26423544550
5	-18470170670	55735427934	24838029904	-18469090086	55735826511	24838016147

TABLE 5 The comparison of the programming and the value of an instance

Point name	Calculate data From arranges	Instance data	Δ
PYX	-9.30886	-9.30886	0
PYY	26.01370	26.01370	0
PYZ	12.29813	12.29813	0
XZX	0.516831	0.516831	0
XZY	-1.218477	-1.218477	0
XZZ	3.506988	3.506988	0
CDBM	-4.2714819	-4.2714819	0

6 Conclusion

Access is one of the Office suites, which is a very powerful tool and easily used in data processing. This paper based on the principle of Bursa-Wolf coordinate transformation model and introduced a self-made simulation program which made full use of the feature that Access VBA can easily link its database. This simulation program has been verified to meet the needs of production by instances. For this, it's easy for non-computer professionals to write a simulation program which is low-cost, convenient, fast, and accurate so that problems in production can be better solved.

References

- [1] Yingqi Guo, Bin Tang, Qiujiang Zhang 2012 *Geodesy and Geodynamics* 32(3) 125-128
- [2] Shi Y M 2008 *Modern geodetic control the measuring* Surveying and Mapping Press: Beijing pp 35-45
- [3] Hapgood M A 1992 *Planetary and Space Science* 40(5) 711-717
- [4] Earl F. Burkholder 2001 *Journal of Surveying Engineering* 127(4) 143-156
- [5] Sprinsky W H 2002 *Journal of Surveying Engineering* 128(4) 200-209
- [6] Jili Yao, Yufei Xu 2007 *Wei Xiao Geo-Spatial Information Science* 10(3) 173-176
- [7] Wenli Wang,Chuanlu Cheng, Junying Chen 2010 *Surveying and Mapping* 35(5) 37-39
- [8] Yuzhong. Shen, Yi Chen, Dehua Zheng 2006 *Journal of Geodesy* 80(5) 233-239
- [9] Li Z H, Huang J S 2005 *GPS measurements and data processing* Surveying and Mapping Press: Beijing pp.167-169
- [10] Hao Chen, Jian Li, Huaxian Yang 2012 *Surveying and Mapping* 10 48-50, 54
- [11] Jixian Wu, Haiyan Du, Yaowen Zhang 2008 *Surveying and Mapping* 33(5) 73-74
- [12] Rey-Jer You, Hwa-Wei Hwang 2006 *Journal of Surveying Engineering* 132(2) 64-70
- [13] Carlson R 2007 *Access Advisor* 15(5) 42-44
- [14] Groh M 2006 *Advisor Guide to Microsoft Access* 14(7) 61-67
- [15] Korn D R 2012 *Modern Machine Shop* 85(6) 108, 110-111, 113-114
- [16] Stockman J 2005 *Access Advisor* 13(3) 38, 40-42
- [17] Stockman J 2006 *Advisor Guide to Microsoft Access* 14(3) 34-39
- [18] Stockman J 2006 *Advisor Guide to Microsoft Access* 14(4) 36-42
- [19] Yajing Liu Shanjun Mao, Dazhi Guo 2006 *Hunan University of Science and Technology (Natural Science)* 21(3) 61-64
- [20] Yanhua Wang, Sherong Hu, Chengshuai Sun 2012 *Surveying and Spatial Information Technology* 35(5) 211-21

Authors



Yusheng Gong

Current position, grades: lecturer of School of Civil Engineering, University of Science and Technology Liaoning
University studies: master's degree from Liaoning Technical University
Scientific interest: majors in GNSS and research and application of Engineering Survey.

New approaches for link prediction in temporal social networks

Nahla Mohamed Ahmed^{1, 2}, Ling Chen^{1, 3*}

¹ College of Information Engineering, Yangzhou University, Yangzhou China, 225009

² College of Mathematical Sciences, Khartoum University, Khartoum, Sudan, 115547

³ State Key Lab of Novel Software Tech, Nanjing University, Nanjing China, 210093

Received 1 March 2014, www.tsi.lv

Abstract

Link prediction in social networks has attracted increasing attention from various domains such as sociology, anthropology, information science, and computer sciences. In this work, efficient approaches to predict potential link in temporal social networks are presented. One approach is based on reduced static graph using a modified reduced adjacency matrix to reflect the frequency of each link. Another approach is based on indices integration, and exploits both the temporal and topological information. The approach integrates the indices in all the time steps, which reflect the topological information, and uses a damping factor to emphasize the importance of more recent links. Experimental results on real datasets show that our approaches can efficiently predict future links in temporal social networks, and can achieve higher quality results than traditional methods.

Keywords: Social networks, temporal networks, link prediction

1 Introduction

Link prediction is an important task in social network analysis. It detects the hidden links from the observed part of the network, or predicts the future links given the current structure of the network. Link prediction has several applications in social network analysis [1-5].

Since relations among social members continuously change over time, links in real world social networks are varying and evolving constantly. Recently, approaches have been advanced to detect potential or future links in such temporal social networks.

Some of such methods are based on the analysis of the topological features of the network [6-8]. H. Kim et al. [8] presented a method to predict future network topology using the nodes' centrality, which can identify the important nodes in the future. Machine learning strategies are also exploited in temporal network link prediction methods [9-12]. Manisha Pujari et al. [9] applied a supervised rank aggregation method for link prediction in temporal complex networks. Some methods for link prediction on temporal network are based on probabilistic model [13-18]. S. Steve Hanneke et al. [13] proposed a family of statistical models for temporal social network link prediction by extending the exponential random graph model. However, such probabilistic model requires a predefined distribution of link appearance, which is difficult to know in advance for a given temporal network.

In this work, we present efficient approaches to predict potential link in temporal social networks. One approach is based on reduced static graph by using a modified reduced adjacency matrix to reflect the frequency of each link. Another approach integrates the similarity indices of the nodes to exploit both the temporal and topological information. Experimental results show that our methods can obtain higher quality results of link prediction in temporal social networks.

2 Reduced static graph approach

2.1 TRADITIONAL REDUCED STATIC GRAPH APPROACH

We use an undirected binary graph to represent the network. Let $V = [v_1, v_2, \dots, v_N]$ be a set of vertices, G_1, G_2, \dots, G_T be a sequence of graphs on V at time steps $t = 1, 2, \dots, T$. Define symmetric matrices A_1, A_2, \dots, A_T as the adjacency matrices of graphs G_1, G_2, \dots, G_T respectively. The binary value of $A_t(i, j)$ indicates the existence of an edge between nodes i and j , $i, j = 1, 2, \dots, N$, during the time period t , $t = 1, 2, \dots, T$. Given such graph series, the goal of link prediction in temporal social network is to predict the occurrence probabilities of edges at time $T+1$.

In the traditional algorithms for link prediction in temporal networks, the networks in time series

* Corresponding author - Tel: +15805271163; fax: +86-514-87887937; E-mail: zylchen@163.com

G_1, G_2, \dots, G_T are reduced into a single weighted graph $G_{1,T}$ represented by a reduced adjacency matrix $\tilde{A}_{1,T}$:

$$\tilde{A}_{1,T}(i, j) = \begin{cases} 1 & \text{if } \sum_{t=1}^T A_t(i, j) \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In the second step, a static network link prediction method is conducted on the reduced static graph $\tilde{G}_{1,T}$, and the result on $\tilde{G}_{1,T}$ is output as the solution of temporal link prediction on G_{T+1} .

2.2 IMPROVED REDUCED STATIC GRAPH APPROACH

The traditional reduced static graph methods for temporal network link prediction have drawbacks of missing some important information. First, since the reduced adjacency matrix $\tilde{A}_{1,T}$ defined in (1) is a binary one, it neglects the frequency of the links in the graphs G_1, G_2, \dots, G_T . It is obvious that if a link occurs more frequently in the temporal network, it will have higher probability to appear in the future. Moreover, the reduced adjacency matrix $\tilde{A}_{1,T}$ ignores the time information, which is also important in link prediction for temporal networks. Since recent links in the network should have more importance in predicting the future links, they should have larger weights in calculating the similarity indexes.

A common feature of existing link prediction approaches under the static graph representation is that the occurrence probability of a link (i, j) is solely determined by other links related to it disregarding the temporal information. The Common Neighbour, Jaccard, Adamic/Adar indices rely on the number of occurrences of link pairs of the form $((i, k), (k, j))$, and ignore previous occurrences of link (i, j) itself. This is the main reason why some predictors based on Common Neighbour index fail to have a high performance in temporal networks with repeated links.

In our model proposed, we first add a self-loop (an edge from node i to j to each node in the graph representation of the social network. Thus if link (i, j) has appeared previously, the link occurrence probability of (i, j) exploits the occurrences of two link pairs $((i, i), (i, j))$, and $((i, j), (j, j))$. In addition, a damping factor is used to assign more importance to the more recent topological information. Therefore, the reduced adjacency matrix $A_{1,T}^*$ is defined as:

$$A_{1,T}^*(i, j) = \begin{cases} 1 & \text{if } i = j \\ \sum_{t=1}^T \delta^{T-(t-1)} A_t(i, j) & \text{otherwise} \end{cases} \quad (2)$$

Here, $0 < \delta < 1$ is a damping factor. The base fact we considered in this model is that recent links have more accurate information than the old ones. In this model, the weight of each edge is reduced by the damping factor at each time. From (2), we also can see that value of $A_{1,T}^*(i, j)$ reflects the frequency of the appearance of link (i, j) . Since more frequently appeared links have higher probability to emerge in the future, and existing links have higher probability to appear in the future than the missing links, the reduced static graph with adjacent matrix $A_{1,T}^*$ is more reliable for temporal link prediction.

In our modified method, we take both the frequency and the time of the link appearance into consideration, the more frequently and recently a link (i, j) appears the higher weight is assigned to it according to (2).

As we will see later from the experimental results, our approach demonstrates higher performance and obtains more accurate results on future link prediction.

3 Indices integration algorithm for temporal link prediction

3.1 FRAMEWORK OF THE ALGORITHM

In this section, we present an indices integration method for temporal link prediction. In the method, we first calculate the indices of all the node pairs in the given temporal graphs G_1, G_2, \dots, G_T . Such indices can be a commonly used similarity measurement such as Common Neighbour, Jaccard, Adamic/Adar, Preferential Attachment or Katz. Let the matrices of the indices be S_1, S_2, \dots, S_T for graphs G_1, G_2, \dots, G_T , respectively. Then we integrate these matrices to construct $S_{1,T}$, which is a matrix consisting of the indices of probabilities for future links.

$$S_{1,T} = \sum_{t=1}^T \delta^{T-(t-1)} S_t \quad (3)$$

In (3), $0 < \delta < 1$ is a damping factor, which is used to assign more importance to the more recent topological information. The probability matrix $S_{1,T}$ carries both time and topological information.

To take the previously appeared links into consideration, in calculation of the indices for each graph G_t , we use the augmented adjacency matrix A_t^* instead of the traditional adjacency matrix A_t . The augmented adjacency matrix A_t^* is defined as:

$$A_t^*(i, j) = \begin{cases} 1 & \\ \text{number of connections between } v_i, v_j & \end{cases} \quad (4)$$

Figure 1 shows the framework of our indices integration algorithm for temporal link prediction.

```

Algorithm (Indices Integration Algorithm for
Temporal Link Prediction)
Input:
  A1*, A2*, ..., AT* : augmented adjacency matrix of
  G1, G2, ..., GT;
  f : A* → S , static graph link prediction algorithm
where A* is an augmented adjacency matrix, and S is a
link occurrence probability score matrix;
  δ : the damping factor 0 < δ < 1
Output:
  S1,T : S1,T(i, j) gives the probability score of edge
(i, j) at time T + 1;
begin
  Initialize a zero N*N matrix S1,T;
  Compute static graph link prediction matrices
  For t=1 to T do
    St = f(At*);
  Estimate the time series link prediction model
  For t=1 to T do
    S1,T = S1,T + δT-(t-1)St
  Output(S1,T)
end
    
```

FIGURE 1 Indices Integration Algorithm for Temporal Link Prediction

3.2 COMPUTATION OF THE INDICES

Since the indices of Common Neighbour, Jaccard, Adamic Adar and Preferential Attachment only consider the direct neighbours for each node, they require computation time less than $O(N^2)$. However, since Katz index considers all existing paths in the graph, its computation time cost is quite expensive. In this section, we provide an efficient way for calculating the Katz index of graphs $G_t, t=1,2,\dots,T$.

The Katz index is defined as follows:

$$S^{KZ} = (I - \beta A)^{-1} - I = \sum_{l=1}^{\infty} \beta^l A^l . \tag{5}$$

Here, A is the adjacency matrix, and $\beta \in (0,1)$ is a factor for emphasising the importance of the short links. S^{KZ} is the matrix consisting of the indices of the node pairs. From (5), we can see that the key challenge in the computation of Katz index is the matrix inversion, which requires $O(N^3)$ time. It is computationally prohibitive for large scale networks. To address this challenge, we

approximate the elements in $Q = (I - \beta A)^{-1} - I$ by rewriting the equation of Q as

$$Q = (I - \beta A)^{-1} - I \approx \sum_{l=1}^{l_m} \beta^l A^l . \tag{6}$$

Here, l_m is a positive integer, $A^l(i, j)$ is the number of paths of length l between nodes i and j . Since the long paths have less influence on the Katz index, we only consider the paths of lengths less than l_m .

In general, the temporal social network evolves smoothly over time, the corresponding sequence of adjacency matrices will not change rapidly. In the other words, the difference matrix $\Delta_t = A_{t+1} - A_t$ is likely to be a sparse one. Under this consideration, we are able to make a quick calculation of Q_{t+1} based on the second-order approximation:

$$Q_{t+1} = Q_t + \beta \Delta_t + \beta^2 (\Delta_t A_t + A_t \Delta_t + \Delta_t^2) . \tag{7}$$

Equivalently,

$$S_{t+1}^{KZ} = S_t^{KZ} + \beta \Delta_t + \beta^2 (\Delta_t A_t + A_t \Delta_t + \Delta_t^2) \tag{8}$$

Since Δ_t has few non-zero elements, terms $\Delta_t A_t, \Delta_t \Delta_t$ and Δ_t^2 in (8) are sparse matrix multiplications. By exploiting the sparseness of matrix Δ_t , we can compute the value of Q_{t+1} very efficiently using incremental proximate updating [22]. As a result, we can compute S_{t+1}^{KZ} in $O(n_t)$ time, where n_t is the number of non-zero elements in Δ_t .

Experiments

To evaluate our proposed methods for link prediction in temporal social network, we test them by a series of experiments on several temporal social networks. All the experiments are performed on a Pentium IV computer running Windows XP, with 1.7G memory, and using VC++ 6.0. First, we introduce the 6 datasets we used in the experiments, and introduce the experimental setup.

4.1 THE TEST DATASETS

4.1.1 Enron Email dataset

Enron Email dataset represents an email communication network of Enron employees during 38 months from May 1999 to June 2002. In the network, nodes represent employees. If there has been at least one email communication between two employees, an edge will link their nodes in the network. We performed the link prediction analysis on the monthly email graphs.

4.1.2 Infectious SocioPatterns (ISP) dataset

Infectious SocioPatterns dataset represents a network reflecting human mobility. It contains the daily cumulated networks represented in the Infectious SocioPatterns visualization system. The nodes represent visitors of the Science Gallery, while the edges represent close-range face-to-face proximity during 20 seconds interval between the concerned visitors.

After pre-processing, we choose the dataset of May first, which includes 8 active hours partitioned into 93 periods. Length of each of period is 5 minutes, therefore, we perform the link prediction analysis on periodically 5 minutes graphs.

4.1.3 Autonomous Systems (AS) dataset

Autonomous Systems dataset represents online communication network of who-talks-to-whom. It was collected by the University of Oregon Route Views Project - Online Data and Reports. The graph consists of routers in the Internet, and is organized into sub-graphs called Autonomous Systems (AS). The nodes represent ASs, and edges represent AS exchanging traffic flows between corresponding nodes. The dataset contains 733 daily instances which span an interval of 785 days from November 8, 1997 to January 2, 2000.

After pre-processing, we chose the first 30 days in the given data, and 470 most active Autonomous Systems. The link prediction algorithm is performed on daily AS communication graphs.

4.1.4 Nodobo dataset

Nodobo dataset represents mobile phone calls network of high-school students, from September 2010 to February 2011. The nodes represent students, and links represents phone calls between corresponding students.

After pre-processing, we choose 61 days starting from the first of October to the 30th of November. The link prediction algorithm is performed on daily phone call graphs.

4.1.5 Manufacturing Emails (ME) dataset

Manufacturing Emails dataset is the internal email communication network between employees of a mid-sized manufacturing company. The nodes represent employees, and an edge between two nodes represent their email communication.

4.1.6 Southern Women (SW) dataset (<http://konect.uni-koblenz.de/networks/opsahl-southernwomen>)

Southern Women dataset shows the participations of 18 women in 14 social events over a nine-month period. The data was collected in the southern United States of America in the 1930s. Originally, this dataset represents a

bipartite network where each edge connects a woman with the event she participates in. We transform this dataset to unipartite temporal networks where the nodes represent women, edges between two nodes in the t -th graph shows that these two women participate in the same event at time t .

Table 1 shows the main features of the 6 datasets, including number of nodes (#Nodes), total number of links (#Links), total number of unique links(# Unique Links), by which repeated links between the same node pair are counted only once, length of the time series sequence (T), total number of unique links at each time step (#Unique Links-T).

TABLE 1 The main features of the datasets

Datasets	#Nodes	#Links	#Unique Links	T	#Unique Links-T
Enron	145	26092	1101	38	3960
ISP	200	5943	714	93	1733
AS	470	92348	1842	30	46174
Nodobo	395	5453	453	61	2016
ME	167	82926	3250	117	29847
SW	14	214	66	18	214

4.2 EXPERIMENT SETUP

For every dataset, we take $T+1$ snapshot graphs G_1, G_2, \dots, G_{T+1} . In each time step $t, t=1, 2, \dots, T$, we use the first t graphs, G_1, G_2, \dots, G_t , to test static and time series link prediction methods to predict links in G_{t+1} . In static graph representation tests, we combine these t graphs to build $G_{1,t}$ as shown in Section 3.

We test our methods based on reduced static graph and on indices integration, and compare the quality of their results with the traditional methods based on reduced static graph. In those traditional methods, we use the similarity measurements of Common Neighbour, Jacard, Adamic/Adar, Katz, and Preferential Attachment; we denote them as CN, JC, AA, KZ and PA respectively. We also test the methods based on the modified reduced adjacency matrix $A_{1,T}^*$ defined in (2) using similarity measurements of Common Neighbour, Jacard, and Adamic/Adar, we denote them as CN*, JC*, AA*, respectively. For the methods based on indices integration, we use both adjacent matrix A_t and augmented adjacency matrix A_t^* . When using adjacent matrix A_t , we use the similarity measurements of Katz and Preferential Attachment, the algorithms are denoted as TKZ, TPA respectively. When using augmented adjacency matrix A_t^* , we use the similarity measurements of Common Neighbour, Jacard, and Adamic/Adar, the algorithms are denoted as TCN*, TJC* and TAA* respectively. Table 2 lists the names of the algorithms, along with their reduction methods and adjacent matrixes used.

TABLE 2 Link prediction methods tested

Reduction method	Adjacent matrix	Common neighbour	Preferential Attachment	Jaccard	Adamic/Adar	Katz
Reduced static graph	$\tilde{A}_{i,T}$ defined in (1)	CN	PA	JC	AA	KZ
	$A_{i,T}^*$ defined in (2)	CN*	-	JC*	AA*	-
Indices Integration	A_i	-	TPA	-	-	TKZ
	A_i^* defined in (4)	TCN*	-	TJC*	TAA*	-

We use AUC (Area Under Curve) scores to evaluate the quality of the results by the algorithms tested. After the algorithms calculating and ranking the similarities of the all the node pairs, which represent all the existent and the non-existent links, the AUC value can be interpreted as the probability that a randomly chosen existent link is given a higher score than a randomly chosen non-existent link. At each time we randomly pick an existent link and a non-existent link to compare their scores, if among n independent comparisons, there are n' times the existent link having a higher score and n'' times they have the same score, the AUC value is

$$AUC = (n' + 0.5n'') / n \tag{9}$$

In general, a larger AUC value indicates a higher performance, hence, AUC value of the perfect result is 1.0, while AUC of the result by a random predictor is 0.5.

4.3 EXPERIMENTAL RESULTS AND ANALYSIS

4.3.1 Tests of the methods based on reduced static graph

First we compare the performance of the reduced static graph methods CN, JC and AA using traditional reduced adjacency matrix $\tilde{A}_{i,T}$ defined in (1) with the methods CN*, JC* and AA* using the modified reduced adjacency matrix $A_{i,T}^*$ defined in (2). Figure 2 shows the AUC values of the results by CN, JC and AA using solid lines, and the AUC values of the results by CN*, JC* and AA* using dash lines. The figure shows a high improvement of AUCs by methods CN*, JC* and AA* implementing modified reduced adjacency matrix $A_{i,T}^*$. The reason for CN*, JC* and AA* getting higher quality results is that the modified reduced adjacency matrix $A_{i,T}^*$ integrates the information of both the frequency and the time of the link appearance. Since the network link connections accumulate over time, the probability of repeated links increases gradually. This explains why the solid lines and dashed lines in Figures 2(A) and 2(B) are close to each other when t is small and run far away when t becomes larger. Figure 2 clearly shows that methods CN*, JC* and AA* can achieve much better performance than other methods.

Table 3 illustrates the average AUC scores of the results by the methods shown in Figure 2. It can be seen from Table 3 that method JC* achieves the best performance on Enron, ISP and Manufacturing Emails datasets, while AA* and CN* obtain the best AUC values on AS and Nodobo datasets respectively. We also can see from the table that methods CN*, JC* and AA*, which use the modified reduced adjacency matrix $A_{i,T}^*$, have higher AUC values than all the other methods on the datasets.

Figure 3 compares the average AUC values by the methods CN*, JC* and AA* with those by PA and KZ. From the figure, we can see that methods CN*, JC* and AA* have better performances than PA, and even KZ. Although method KZ has a deep consideration on common paths, methods CN*, JC*, AA* get even better results.

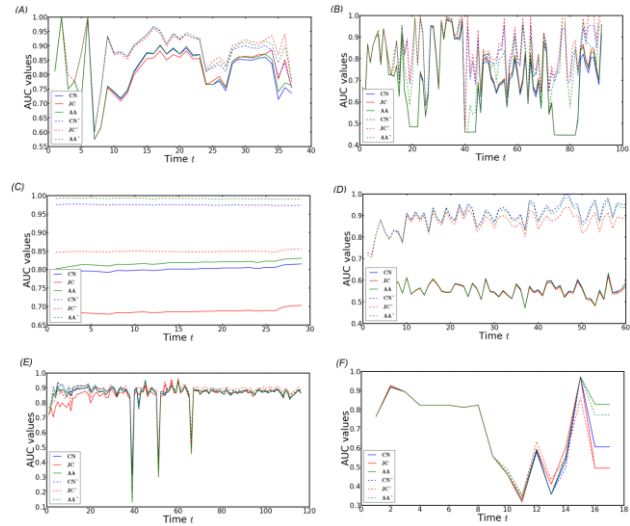


FIGURE 2 Testing results by methods CN, JC, AA, CN*, JC* and AA* on datasets:(A) Enron, (B) ISP, (C) AS, (D) Nodobo, (E) ME, (F) SW

Table 4 presents the average AUC values of CN*, JC*, AA*, PA and KZ on the 6 datasets. It can be seen from Table 4 that method JC* achieves the best performance on Enron and ISP datasets, while AA* and CN* obtain the best AUC values on AS and Nodobo datasets respectively. It is obvious that methods CN*, JC* and AA*, which use modified reduced adjacency matrix $A_{i,T}^*$, have higher AUC values than the other methods on most of the datasets. This demonstrates that using the modified reduced adjacency matrix is helpful for increasing the quality of link prediction in temporal social networks.

TABLE 3 Average AUC values by methods CN, JC, AA, CN*, JC* and AA* on 6 datasets

Algorithm	Enron	ISP	AS	Nodobo	ME	SW
CN	0.8095	0.6966	0.8006	0.5546	0.857	0.6867
JC	0.8133	0.7064	0.6861	0.5505	0.8441	0.682
AA	0.8148	0.7029	0.8174	0.5543	0.8587	0.7153
CN*	0.8735	0.8242	0.9748	0.9004	0.8725	0.6884
JC*	0.8827	0.8423	0.8492	0.868	0.8772	0.6776
AA*	0.8797	0.7969	0.9918	0.8917	0.8748	0.7116

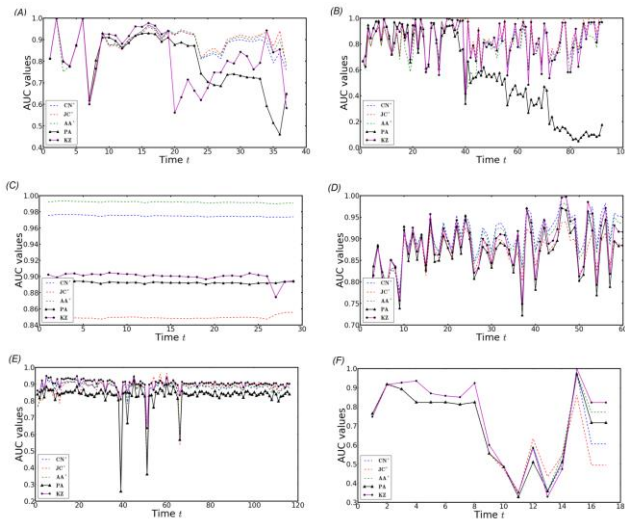


FIGURE 3 Testing results by methods CN*, JC*, AA*, and PA, KZ on datasets: (A) Enron, (B) ISP, (C) AS, (D) Nodobo, (E) ME, (F) SW

TABLE 4 The average of AUC values of CN*, JC*, AA*, PA, and KZ

Algorithm	Enron	ISP	AS	Nodobo	ME	SW
CN*	0.8735	0.8242	0.9748	0.9004	0.8725	0.6884
JC*	0.8827	0.8423	0.8492	0.868	0.8772	0.6776
AA*	0.8797	0.7969	0.9918	0.8917	0.8748	0.7116
PA	0.7993	0.5521	0.8923	0.8673	0.8301	0.6964
KZ	0.8194	0.8305	0.8998	0.8822	0.9019	0.7349

4.3.2 Tests of the methods based on indices integration

We also test our indices integration methods TCN*, TJC*, TAA*, TPA and TKZ on the 6 data sets. Figures 4 to 9 show the AUC values of the results on the datasets of Enron, ISP, AS, Nodobo, ME and SW respectively. We also compare the AUC values of the results by methods TCN*, TJC*, TAA*, TPA, and TKZ with their corresponding reduced static graph methods CN, JC, AA, PA and KZ. From the figures, we can see that AUC values of the results by methods TCN*, TJC*, TAA*, TPA and TKZ are all higher than their corresponding static graph methods on all the data sets. For instance, a surprising result we achieved is the amazing improvement of AUC value by TPA over PA on Infectious SocioPatterns dataset. When t is greater than 40, AUC value TPA increases rapidly, and becomes 3 to 10 times higher than that of PA. Our methods based on indices integration can achieve higher quality results because they consider both temporal and topological information, while static graphs methods lose time serial information. In our methods, temporal representation provides better improvement on the results of link prediction.

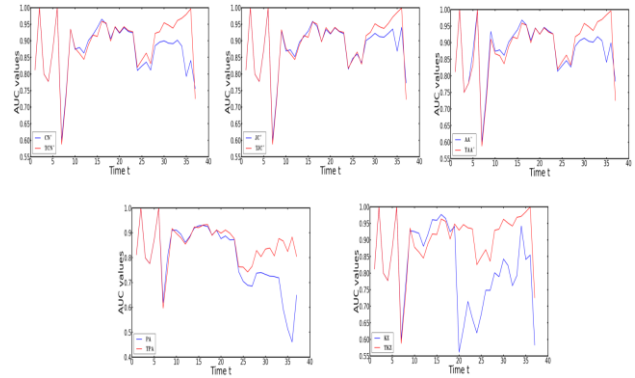


FIGURE 4 Testing results by methods based on indices integration on Enron dataset

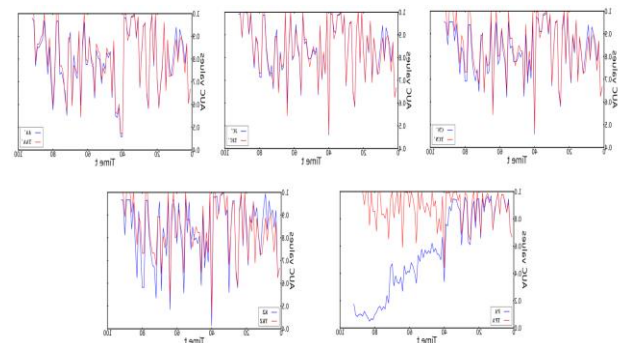


FIGURE 5 Testing results by methods based on indices integration on Infectious SocioPatterns dataset

Table 5 shows the experimental results by our indices integration based methods TCN*, TJC*, TAA*, TPA, TKZ and their counterparts based on static graph. The reported values in the table are the average AUC scores of tests at 60 time steps on the 6 datasets. It shows that indices integration based Katz method, TKZ, achieves AUC values 0.9845, 0.9014, 0.9359 and 0.7965 on Enron, Nodobo, and Manufacturing Emails and South Women datasets respectively. These are the best AUC values among all the methods on those datasets. The indices integration based Preferential Attachment method, TPA, has the highest AUC score 0.8909 for Infectious SocioPatterns dataset, while the indices integration based Jacard method, TJC*, obtains the highest AUC score 0.9932 on Autonomous Systems dataset.

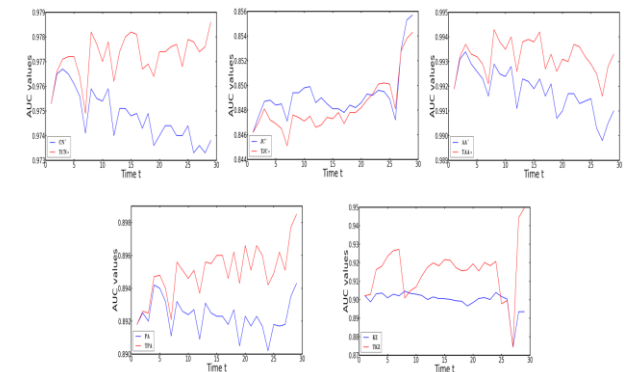


FIGURE 6 Testing results by methods based on indices integration on Autonomous systems dataset

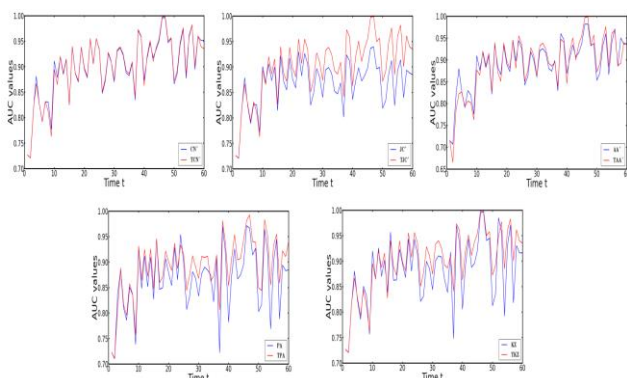


FIGURE 7 Testing results by methods based on indices integration on Nodobo dataset

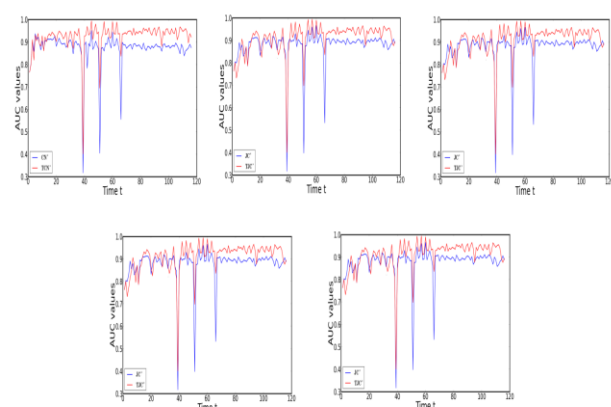


FIGURE 8 Testing results by methods based on indices integration on Man.Email dataset

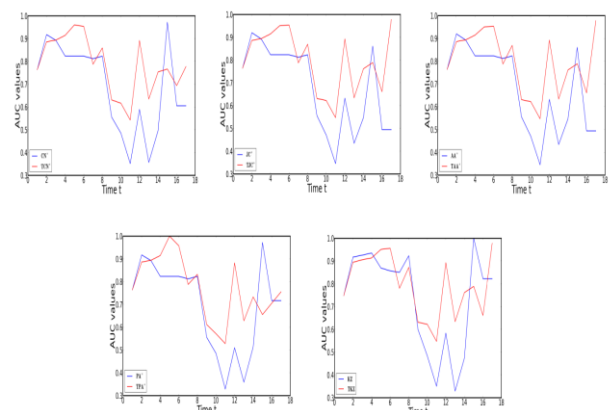


FIGURE 9 Testing results by methods based on indices integration on SW dataset

It is worth mentioning that the best AUC values we achieved in all experiments on Enron, ISP, AS, Nodobo and Manufacturing Email datasets are 0.8945, 0.8909, 0.9932, 0.9014 and 0.9539 respectively. These significantly high AUC values are all obtained by our indices integration based methods. By using the augmented adjacency matrix, we can improve the quality of the results of link prediction in temporal social networks. Moreover, we reach even better improvement by the methods using indices integration, which can

integrate the temporal and the topological information of the temporal social networks.

TABLE 5 Average AUC values for static and time series graph representations

Algorithm	Enron	ISP	AS	Nodobo	ME	SW
CN*	0.8735	0.8242	0.9748	0.9004	0.8725	0.6884
JC*	0.8827	0.8423	0.8492	0.868	0.8772	0.6776
AA*	0.8797	0.7969	0.9918	0.8917	0.8748	0.7116
PA	0.7993	0.5521	0.8923	0.8673	0.8301	0.6964
KZ	0.8194	0.8305	0.8998	0.8822	0.9019	0.7349
TCN*	0.8914	0.8422	0.9772	0.9008	0.9253	0.7843
TJC*	0.8902	0.8424	0.8483	0.9005	0.9092	0.7964
TAA*	0.8877	0.8034	0.9932	0.8924	0.9251	0.7775
TPA	0.8524	0.8909	0.895	0.8938	0.8749	0.7712
TKZ	0.8945	0.8423	0.9155	0.9014	0.9359	0.7965

5 Conclusions

We investigate the problem of link prediction in temporal social networks. In this work, we achieve higher quality link prediction results by providing larger weights for frequently occurred links. We also present a time series model that exploit temporal information on evolving social network for link prediction. We advanced a method based on indices integration which exploits both the temporal and topological information of the temporal networks. To take the previously appeared links into consideration, we define an augmented adjacency matrix in calculating the indices at each time step. We provide a fast algorithm for efficiently calculating the indices involving matrix computation such as Katz index. We conduct extensive experimental evaluation of our methods on benchmark datasets. Experimental results show that our methods can obtain higher quality results of link prediction in temporal social networks.

Acknowledgments

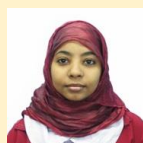
This research was supported in part by the National Natural Science Foundation of China (61379066, 61070047, 61070133), Natural Science Foundation of Jiangsu Province (BK21010134, BK20130452) Natural Science Foundation of Education Department of Jiangsu Province (09KJB20013).

References

- [1] Lü L Y, Zhou T 2011 *Physica A* **390** 1150–1170
- [2] Ahmed H S, Faouzi B M, Caelen J 2013 Detection and classification of the behaviour of people in an intelligent building by camera *International Journal on Smart Sensing and Intelligent Systems* **6**(4) 1317-1342
- [3] Yang L T, Wang S, Jiang H 2011 Cyclic Temporal Network Density and its impact on Information Diffusion for Delay Tolerant Networks *International Journal on Smart Sensing and Intelligent Systems* **4**(1) 35-52
- [4] Liu Z H, Ma J F, Zeng Y 2013 Secrecy transfer for sensor networks: from random graphs to secure random geometric graphs *International Journal on Smart Sensing and Intelligent Systems* **6**(1) 77-94

- [5] Jia X, Xin F, Chuan W R 2013 Adaptive spray routing for opportunistic networks *International Journal on Smart Sensing and Intelligent Systems* 6(1) 95-119
- [6] Purnamrita S, Chakrabarti D, Jordan M 2012 Nonparametric link prediction in dynamic networks *arXiv preprint arXiv 1206-6394*
- [7] Murata T, Moriyasu S 2008 Link Prediction based on Structural Properties of Online Social Networks *New Generation Computing* 26(3) 245-257
- [8] Kim H, Tang J, Anderson R, Mascolo C 2012 Centrality prediction in dynamic human contact networks *Computer Networks* 56 983-996
- [9] Pujari M, Kanawati R 2012 *Supervised Rank Aggregation Approach for Link Prediction in Complex Networks* Companion, Lyon, France pp. 1189-1196 April 16–20, 2012
- [10] Vu D Q, Asuncion A U, Hunter D R, Smyth P 2011 Continuous-time regression models for longitudinal networks *Advances in Neural Information Processing Systems 24: Proceedings of the 25th Annual Conference on Neural Information Processing Systems* 1-9
- [11] Bringmann B, Berlingerio M, Bonchi F, Gionis A 2010 Learning and Predicting the Evolution of Social Networks *IEEE Intelligent Systems* 26-34
- [12] O'Madadhain J, Hutchins J, Smyth P 2005 Prediction and ranking algorithms for event-based network data *In Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 23-30
- [13] Hanneke S, Fu W J, Xing E P 2010 Discrete temporal models of social Networks *Electronic Journal of Statistics* 4 585–605
- [14] Ji Liu, Denga G 2009 Link prediction in a user_object network based on time-weighted resource allocation *Physica A* 388 3643-3650
- [15] Gao S, Denoyer L, Gallinari P 2011 Temporal Link Prediction by Integrating Content and Structure Information *CIKM'11, Glasgow, Scotland, UK* 1169-1174
- [16] Potgieter A, April K, Cooke R, Osunmakinde I O 2006 Temporality in link prediction: Understanding social complexity *Journal of Transactions on Engineering Management* 11(1) 83-96
- [17] Hu F Y, Wong H S 2013 Labelling of Human Motion Based on CBGA and Probabilistic Model *International Journal on Smart Sensing and Intelligent Systems* 6(2) 583-609
- [18] Jianhan Zhu, Jun Hong, Hughes J G 2002 Using Markov Chains for Link prediction in Adaptive Web Sites D. Bustard, W. Liu, and R. Sterritt (Eds.): *Soft-Ware 2002, LNCS 2311* 60–73

Authors



Nahla Mohamed Ahmed, born in July, 14, 1987, Khartoum, Sudan

Current position, grades: PhD student, MS

University studies: Khartoum University, Khartoum, Sudan

Scientific interest: Mathematics

Experience: She got her master degree in mathematics in Institute of Mathematical Sciences, Cape town, South Africa. She is currently a lecturer in College of Mathematical Sciences, Khartoum University, Sudan. Her research interest is in complex network analysis.



Ling Chen, born in September, 10, 1951, Jiangsu, China

Current position, grades: Professor

University studies: Yangzhou University, China

Scientific interest: Computer Science

Publications: Has published more than 200 journal and conference papers in computer science.

Experience: He is currently professor of computer science, and the dean of Information Technology College, Yangzhou University, Jiangsu Province, P.R. China. He is also the chair of Yangzhou Computer Society, vice chair of Jiangsu Computer Society, senior member of the Chinese Computer Society, member of IEEE Computer Society. His research interests include parallel processing, computer architecture, algorithms of optimization.

Modelling study of p-xylene oxidative side-reaction based on Bayesian filtering

Liu Ranran^{1, 2}, Li Zhengming^{1*}

¹ School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

² Changzhou Liu Guojun Vocational Technology College, Changzhou 213004, China

Received 1 March 2014, www.tsi.lv

Abstract

The combustion loss model of HAC and PX was established based on Bayesian filtering. Moreover, this neural network model regarded the main adjustable process parameters (residence time, the concentration of Co, Mn, Br, reaction temperature and solvent ratio) as the independent variables and the total content of CO and CO₂ as the dependent variable. The simulation results show that the network is well performed. The effects of process operating parameters on HAC and PX combustion loss are analysed based on the model. Wherein, the effects of residence time, reaction temperature on the combustion loss are bigger, while the effects of other factors are relatively small.

Keywords: Mathematical model, PX oxidative side-reaction, Bayesian filtering, soft sensor

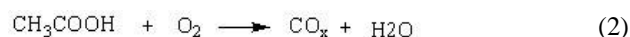
1 Introduction

PTA (PTA) is an important raw material of polyester fiber and plastic. Polyethylene terephthalate (PET) can be obtained from PTA and polyethylene glycol through a directly polycondensation process. As thermoplastic polyester, PET has extensive application in many areas. Meanwhile, as the largest output of dicarboxylic acid, terephthalic acid (TA)'s demand is great in the world. TA can be obtained from PX through a Liquid-phase catalytic oxidation action in AMOCO process, in which catalyst is Co and Mn, promoter is Br, and solvent is acetic acid (HAc) [1-2]. In addition to the main response, PX oxidative is a complex chemical reaction accompanied by a large number of side-reaction. Burning of HAc and PX is the principal side-reaction, whose main product is the carbon oxygen compounds, mainly as CO and CO₂ [3-5]. Besides the production losses of HAc and TA, by-products of HAc and PX oxidative side-reaction can lead to catalyst deactivation. Meanwhile, macromolecule side-reaction formed by the burning of PX can damage the quality of TA and PTA., and limit the mother liquor circulation. Thus not only caused a drop in the quality of products, but also caused an increase in the consumption of catalyst and promoter.

In this paper, a model of PX oxidative side-reaction was studied on the base of the plants in Yang Zi Petrochemistry Company (YPC) to provide basis for the optimization of operation parameters and the reform of production process.

2 Auth mechanism of PX oxidative side-reaction

PX oxidative reaction is in the condition of high temperature, excess oxygen and high-intensity stir. At the same time of main response, PX and HAc will partly combust to carbonic oxide (CO), carbonic dioxide (CO₂) and water according to the following reactive:



Generally, in the industrial operation condition, CO, CO₂ and methyl acetate are the by-products of the HAc's burning which losses HAc about 2% of its circulation. Carbon oxides and other by-products of PX burning make up about 2% of its feedstock. CO_x from the burning of HAc account for 60%, in which contain CO and CO₂ about 60%, while CO_x from the burning of PX is 40%, in which CO and CO₂ is 75% [6].

In the process of PX oxidation, a certain oxidation depth is requested in order to reach some conversion and absorption rate, especially reduce the concentration of 4-CBA. Over oxidation should be controlled because it can make HAc, 4-CBA and its by-products further oxidation to CO₂ and H₂O. The most important factors influent side-oxidation are: reaction temperature (x_1 , °C), solvent ratio (x_2 , Kg.HAc/Kg.PX), Co concentration (x_3 , wt%), Mn concentration (x_4 , wt%), Br concentration (x_5 , wt%) and reaction time (x_6 , s). These operation parameters not

*Corresponding author -Tel: 0511-88790618;
E-mail: lzming@ujs.edu.cn

but control PX oxidation reaction, but also affect HAC and PX side-oxidation reaction [7-8].

3 Bayesian estimation of side-reaction

3.1 BAYESIAN ESTIMATION FIGURES AND TABLES

In this research, the content of CO and CO₂ was thought as a state estimation problem. The state-space model was utilized to describe this problem. Assume that the position state of the target is $(x_t|t=1,\dots,N)$, and the collection for each moment observation is $(z_t|t=1,\dots,N)$. Then the state of the target can be described by the motion equation and observation equation in formula 3:

$$\begin{cases} x_t = F(x_{t-1}) + \omega_t \\ z_t = H(x_t) + \nu_t \end{cases}, \quad (3)$$

where F is the impact parametric equation describing the model input;

H is the observation model, which describes the relationship between the observed quantity and model output;

ω_t is the noise, which is utilized to describe the uncertainty of movement;

ν_t is the observed noise, which is utilized to describe the uncertainty caused by outside interference and the noise of the detecting element.

Bayesian estimation is a state estimation method utilizing state priori distribution and the observation of likelihood function to determine the posterior probability distribution. For a first-order Markov process, assume that the observation of each moment is mutually independent. If the posterior distribution at time t-1 is $p(x_{t-1}|z_{t-1})$, the prior distribution at time t can be expressed as shown in formula 4:

$$p(x_t|z_{t-1}) = \int p(x_{t-1}|z_{t-1})p(x_t|x_{t-1})dx_{t-1}, \quad (4)$$

where $p(x_t|x_{t-1})$ is the transition probability density function, which is determined by F and the probability distribution of the noise $\omega_t(p(\omega_t))$ in motion equation 3. The definition is:

$$p(x_t|x_{t-1}) = \int p(\omega_t)\delta(x_t - F(x_{t-1}))d\omega_t, \quad (5)$$

wherein $\delta(\cdot)$ is the Dirac function. After the prior distribution $p(x_t|z_{t-1})$ is obtained, the state posterior distribution $p(x_t|z_t)$ could be expressed as:

$$p(x_t|z_t) = \frac{p(z_t|x_t)p(x_t|z_{t-1})}{\int p(z_t|x_t)p(x_t|z_{t-1})dx_t}, \quad (6)$$

where $p(z_t|x_t)$ is the observation likelihood function, which is determined by H and the probability distribution of the noise $\nu_t(p(\nu_t))$. The definition is:

$$p(z_t|x_t) = \int p(\nu_t)\delta(z_t - H(x_t))d\nu_t. \quad (7)$$

The above formula 4 and 5 constitute the prediction process of Bayesian estimation; formula 6 and 7 constitute the update process of Bayesian estimation. The two processes are determined by the equation and observation equation of state space model parameters in formula 3, respectively. The tracking of the posterior distribution of state x_t could be achieved through the iterative and recursive solution of above prediction and update process.

3.2 MODELING COX IN TAIL GAS OF SIDE-OXIDATION

The source of our data was the PTA production process data in a chemical plant. The model of paraxylene (PX) oxidation side reaction through soft sensor model provided the basis for production optimization of operating parameters and the transformation of the production process. The main factors that will affect the combustion side reaction are: the reaction temperature (x_1 , °C), solvent ratio (x_2 , Kg. HAC/Kg. PX), the concentration of cobalt catalyst (x_3 , wt%), the concentration of manganese catalyst (x_4 , wt%), the concentration of bromine accelerator (x_5 , wt%) and the residence time (x_6 , S). Regard these main factors after data pre-processing as the input data; the total content of CO and CO₂ (ΦCO_x) of the reactor exhaust gas as the output data y. Generally, 250 sets of data were obtained. 170 sets of data are the training data, and the remaining 80 sets of data are the current running data in the process, which were utilized to test the generalization capability of the model.

Further simulation analysis on the training results and the linear regression analysis of the output and target output of network simulation are shown in figure 1. The figure shows that the correlation coefficient is 0.963, which means that the performance of the network is satisfying.

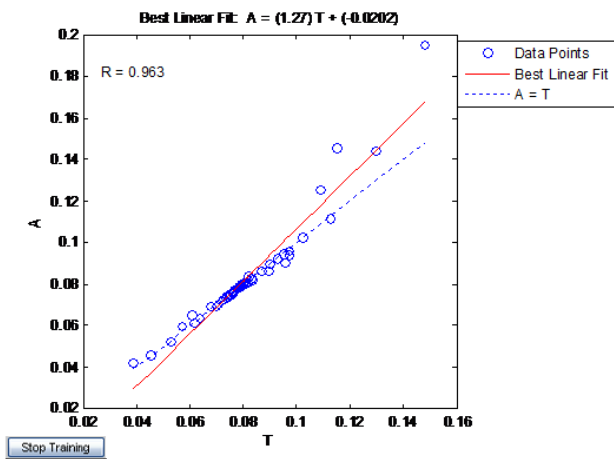


FIGURE 1 The analysis result of network output

3.3 THE COMBUSTION LOSS MODEL OF ACETIC ACID AND XYLENE

Establish the combustion loss model of acetic acid and PX based on the determined generation model of CO_x. The combustion loss model of acetic acid:

$$m_{HAc}^{consume} = \frac{m_{gas} \times x_{CO_x} \times \frac{x_{HAc}}{100} \times \frac{60}{1000}}{2 \times \frac{m_{HAc}^{CO_x}}{100} \times m_{CTA}}$$

The combustion loss model of PX is:

$$m_{PX}^{consume} = \frac{m_{gas} \times x_{CO_x} \times \frac{x_{PX}}{100} \times \frac{106}{1000}}{8 \times \frac{m_{PX}^{CO_x}}{100} \times m_{CTA}}$$

x_{CO_x} : the total content of generated CO and CO₂ of the exhaust gas

x_{HAc} : the percentage of CO_x generated by acetic acid (%)

$m_{HAc}^{CO_x}$: the percentage of CO_x accounting for the combustion product of acetic acid side reaction (%)

m_{gas} : the flow of exhaust gas from the reactor (ft³/hr)

m_{CTA} : CTA product yield (tons / Hr)

The molecular weight of acetic acid: 60 (g / mol)

$m_{HAc}^{consume}$: the amount of acetic acid loss (Kg / ton. CTA)

x_{PX} : the percentage of CO_x generated by PX (%)

$m_{PX}^{CO_x}$: the percentage of CO_x accounting for the combustion product of PX side reaction (%)

The molecular weight of PX: 106 (g / mol)

$m_{PX}^{consume}$: the amount of PX loss (Kg / ton CTA)

From the above text, the value of x_{HAc} , $m_{HAc}^{CO_x}$, x_{PX} and $m_{PX}^{CO_x}$ are known as 60%, 75%, 40% and 60%, respectively; m_{gas} and m_{CTA} were obtained through the actual production data.

4 Impaction to side-oxidation of operating parameters

An impact model of operating parameters and side-oxidation of HAc and PX is presented. The impact of each operating parameters (Co, Mn, Br, reaction time, reaction temperature etc.) to side-oxidation is analysed in the model. Shows in Fig2~Fig6).

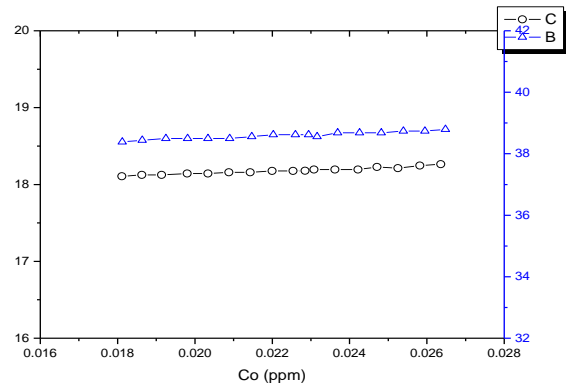


FIGURE 2 Effect of Co concentration on oxidative side-reaction

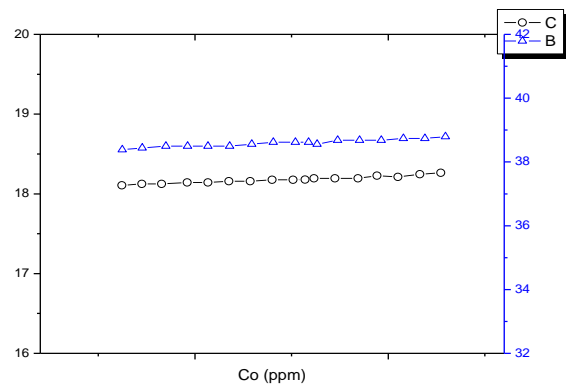


FIGURE 3 Effect of Mn concentration on oxidative side-reaction

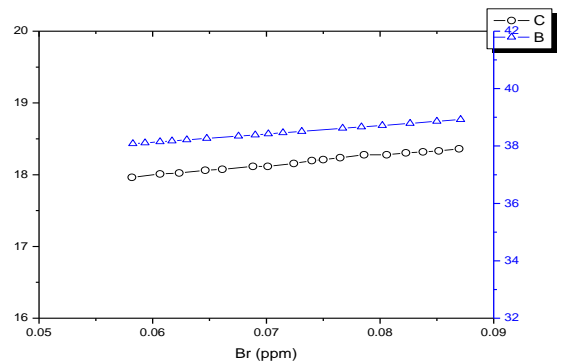


FIGURE 4 Effect of Br concentration on oxidative side-reaction

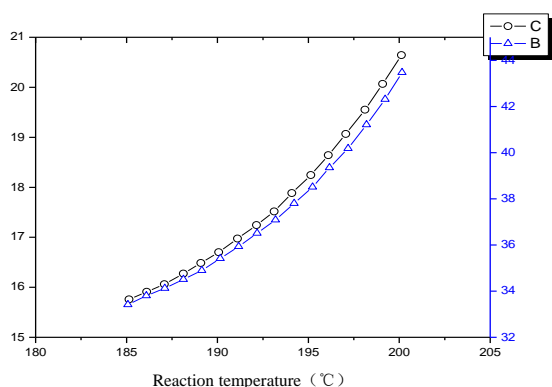


FIGURE 5 Effecton of reaction temperature on oxidative side-reaction

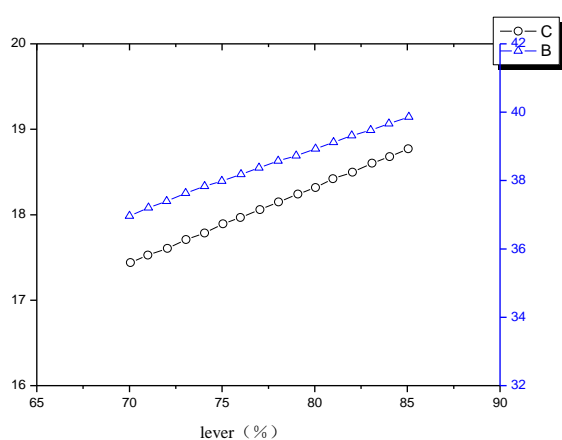


FIGURE 6 Effecton of lever on oxidative side-reaction

From the Figs, it can be seen that the impacton of reaction temperature is greatest, reaction time is greater, and others is less. Quantitative analysis of the impacton can be obtained from the figs, to guide the operating parameters' adjustment and optimization.

5 Conclusions

The model of p-xylene (PX) oxidative side-reaction was proposed in this paper. The Bayesian estimation technique was employed to model the relationship between the main operation conditions (i.e. residence time, reaction temperature and catalyst concentration) and the content of carbon dioxide in the reaction off-gas based on the factors. Furthermore, put forward the combustion expanse model of acetic acid and PX. Impacton of HAc and PX consume was analysed based on the network. Result is that the impacton of reaction temperature is greatest and others are less.

Acknowledgements

The authors would like to thank the financial support provided by Research and Innovation Project for College Graduates of Jiangsu Province of China under Grant CXLX12_0648 and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- [1] Eason G, Noble B, Sneddon I N 1955 On certain integrals of Lipschitz-Hankel type involving products of Bessel functions *Phil. Trans. Roy. Soc. London*, **A247** 529–551
- [2] Maxwell J C 1892 A Treatise on Electricity and Magnetism *Oxford: Clarendon* **3**(2) 68–73
- [3] Jacobs I S, Bean C P 1963 Fine particles, thin films and exchange anisotropy *Magnetism* **3** G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271–350
- [4] Elissa K Title of paper if known *Unpublished*
- [5] Nicole R Title of paper with only first word capitalized *J. Name Stand. Abbrev.* In press
- [6] Yorozu Y, Hirano M, Oka K, Tagawa Y 1987 Electron spectroscopy studies on magneto-optical media and plastic substrate interface *IEEE Transl. J. Magn.* **2** 740–741 Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982
- [7] Young M 1989 *The Technical Writer's Handbook* Mill Valley, CA: University Science

Authors



Liu Ranran, born in May, 24, 1982, Shijiazhuang, Hebei, China

Current position, grades: PhD

University studies: Jiangsu University

Scientific interest: System identification, mathematics modelling.

Publications: 7

Experience: Starting with master period in 2007, I devote myself in the simulation of chemical industry. Especially in the developing of the simulation system of PTA for YPC. In PhD, the main scientific interest are system identification, mathematics modelling, soft sensors and systems simulation.



Li Zhengming, born in 1958, Shuyang, Jiangsu, China

Current position, grades: Professor

University studies: Xian Jiaotong University

Scientific interest: Smart grid and remote monitoring system, computer networks and its application in industrial control, modern information processing techniques.

Publications: more than 100

Experience: Born in 1958, at present hold the office of vice-president in School of Electrical and Information Engineering of Jiangsu University, professor and PhD supervisor. Graduated from Xian Jiaotong university in 1987, obtaining the master degree of control theory and control engineering. Scientific interesting are smart grid, remote monitoring system, computer networks and its application in industrial control and modern information processing techniques. The publications are more than 100, among which more than 20 are included by SCI, EI.

Bipartite graph model for RDF data cleansing

Huang Li^{1, 2*}

¹ College of Computer Science and Technology, Wuhan University of Science and Technology

² Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System

Received 1 March 2014, www.tsi.lv

Abstract

Many systems use RDF to describe information resources and semantic associations between resources. RDF data plays a very important role in advanced information retrieval. Due to diversity and imprecise of resources, duplicates exist in RDF data. The query and retrieval of RDF data are studied by many researchers. However, researchers seldom study RDF data cleansing. In this paper, we focus on RDF data cleansing. According to the features of RDF data, we propose a new approach. This approach combines similarity and connections among resources. First, we introduce an intermediate model, named RDF-Bipartite Graph model, to represent the RDF data. This model improves from Bipartite Statement-Value Graphs model. Then on the proposed model, we design a Subgraph-Extend method, to find the path connecting two nodes. This method detects the minimum subgraph containing two nodes for connect-path finding. It avoids the connect-weight setting in traditional method. Experiments on publication datasets show that the proposed method is efficient in duplicate detection of RDF data, and has high performance and accuracy.

Keywords: RDF, data cleansing, Bipartite Graph

1 Introduction

Revolution of information technology has led to an explosion in the growth of data in digital form. Then this incurs a problem: how to manage and use such large amount of data. One difficulty to manage those data is their diverse sources, for example, the data are managed by different organization and described and stored in different form, although data integration systems and search engines can help users manage them in some sense. However, due to diversity of data resources, there is a large amount of duplicates after data integration.

The semantic associations among resources are as the same important for information querying and retrieval as information resources themselves. Many systems use RDF (Resource Description Framework) data to represent resources and the semantic associations among resources. Due to the existence of duplicates in different resources, duplicate still exist in RDF data integration. Thus RDF data need to be cleaned.

We propose an approach to detect duplicates in RDF data. The duplicate detection method in this approach combines the relationships and similarities among resources. To extract more information of relationships among resources, we improve the Bipartite Statement-Value Graph model for RDF [1]. We call the new intermediate model RDF-Bipartite Graph Model. The relationships are presented as connecting path among data in this model. We also propose a method, Subgraph-Extend method, to find a path between two nodes in RDF-Bipartite Graph. This method utilizes the attributes

of RDF-Bipartite Graph. This method extends subgraph to find the connect path, instead of find shortest path using the sum of weights of edges. Then the resistance of connect-path is used to detect duplicates instead of connect-weight.

The rest of the paper is organized as follows. The RDF-Bipartite Graph model is defined in section 2. Section 3 describes the method of finding connect path among data. In section 4, we report the performance of our approach. Related work is discussed in Section 5. Finally, we conclude this paper in section 6.

Your goal is to simulate the usual appearance of papers in a Journal of the Academy Publisher. We are requesting that you follow these guidelines as closely as possible.

2 Bipartite Graph Model for RDF Data

2.1 RDF DATA

RDF (Resource Description Framework) is proposed by WWW Consortium. It is a language to describe metadata about information resource. RDF statement is a triple. It is consisted of a subject, a predicate and an object. A triple means to assign a value (subject) to one kind property (predicate) of the resource (object). The subject is the resource, which is described, the predicate describes some property, and the object is the value of the property. A set of RDF statements composes a RDF Graph. Figure 1 shows part of RDF graph.

* Corresponding author - Tel: +86-130-988-82821; fax: +86-027-6889-3420; E-mail: huangli82@wust.edu.cn

In Figure 1, we can find out that an RDF graph is not a simple direct graph. There may be more than one edges between two nodes. If we only use a direct graph to represent the RDF data, some information must be missing. In RDF graph, each edge refers to an RDF statement. Every edge connects three nodes. Therefore, the RDF graph is a hypergraph. In a hypergraph, the hyperedges express the statements, and the hypernodes denote subjects, predicates and objects.

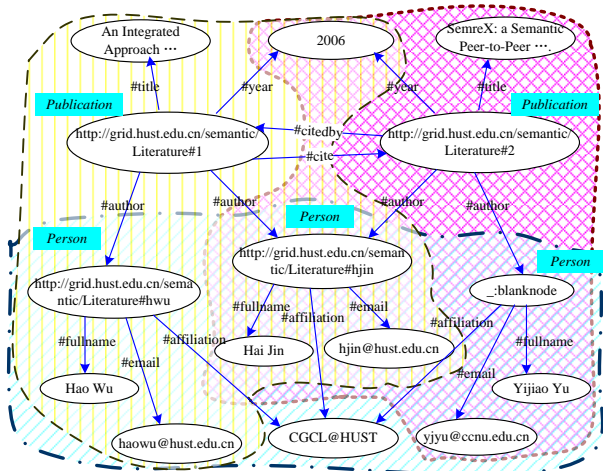


FIGURE 1 Part of RDF graph

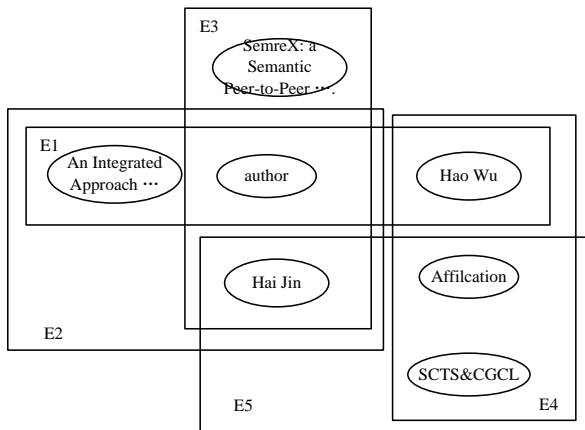


FIGURE 2 Part of hypergraph of Figure 1

Figure 2 shows a part of hypergraph of figure 1. There are seven hypernodes denoted by elliptoid point and five hyperedges: E1, E2, E3, E4 and E5. All hyperedges connect three nodes.

2.2 RDF-BIPARTITE GRAPH MODEL

In [1], Hayes et al. proposed bipartite graphs model to represent RDF data. In the bipartite graphs model, both statement and value are represented by a node in the graph. They transform the RDF graph into a bipartite graph. The two sides of nodes are statement nodes and value nodes. The bipartite graph is a hierarchical graph. Moreover, the entities in the same layer have strong

relationship. All subjects, objects and predicates are called as value statements.

The hierarchy bipartite graph is structured as follows. Because the subject and object of one statement are in one side of bipartite graph, the subject and object are in the same layer, in each statement. The predicate is in the other layer. Moreover, we rule the layer of predicate is higher.

The path between two nodes of same layer in RDF-Bipartite graph is referred the degree of connection between them. If the path is shorter and weightier, the connection is stronger.

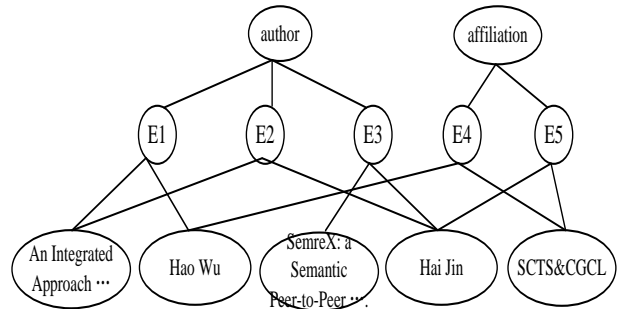


FIGURE 3 Bipartite graph of naive hypergraph in Figure 2

For duplicate detecting, the relationships among duplicates are indirect. Duplicates must be of the same type. The subjects and the objects are of the same type. Therefore, the duplicates are on the same layer. Thus, the path is found in the same layer. That is the nodes in the paths must be in the same and higher layers. Duplicates may exist in value nodes. The statement nodes are the bridges of value nodes. We name the above path, d-path, for distinguishing with other paths in the graph.

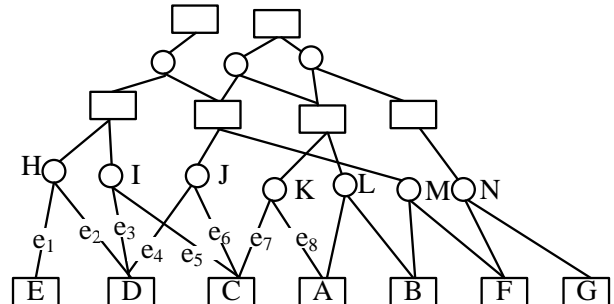


FIGURE 4 Toy RDF-Bipartite graph. The circular points denote statement nodes. And value nodes are denoted by rectangle points.

For example, in figure 4, the paths from 'E' to 'A' are $\{(e_1, e_2, e_3, e_5, e_7, e_8), (e_1, e_2, e_4, e_6, e_7, e_8)\}$. The d-paths connecting two value nodes in the same layer are not unique.

First, we give some definitions.

Definition 1. Connect-weight. Connect-weight is to measure the connective level of path, which connect two nodes. It is the sum of all the edges weights in the path.

Definition 2. Connect-Path. For all paths connecting two nodes, the path which has the biggest connect-weight is called connect-path.

Then the problem changes to find the connect-path between two nodes.

In an RDF-Bipartite graph model, because all RDF statements are consisted of 3-tuple, the degree of statement nodes are three. Moreover, according to the characteristic of RDF-Bipartite graph model, only subject and object of a statement are in the same layer. If a path passes through statement node, it must be preceded by subject and followed by object of the statement, or preceded by object and followed by subject of the statement.

In next section, we will propose a method to find connect-path between two nodes in RDF-Bipartite graph.

3 Connect-Path

3.1 SUBGRAPH-EXTEND

Before describing SE, we will introduce some new concepts used in SE. The subgraph in this method is a special subgraph. This subgraph is consisted of units instead of nodes. There can be many nodes in a unit.

According to the definition of RDF-Bipartite, RDF-Bipartite graph is consisted of 3-tuples. We can take a 3-tuple as a unit in RDF-Bipartite. Figure 5(a) shows a unit in a subgraph of RDF-Bipartite. And Figure 5(b)(c) shows different kinds of subgraphs in RDF-Bipartite.

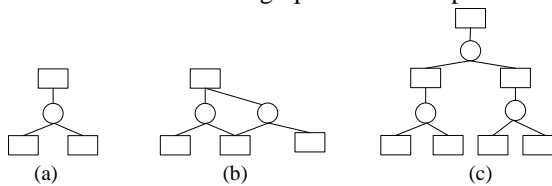


FIGURE 5 Subgraphs of RDF-Bipartite

Fig. 5(b) is consisted of two units, and (c) has three units. In SE, one subgraph is extended by units. Fig 5(b) can be considered as extended from 5(a), and 5(c) could be extended from 5(a) or 5(b).

In fig 5, we can also find that the units in one subgraph overlap with each other. If the units are independent, the subgraph which are consist by these units is not a connective subgraph. This subgraph is to help finding connect-path. And the subgraphs are all connect graph. Thus, in SE, we extend the subgraph by units, which overlap with the exiting units. Obviously, the more two units overlap, the stronger connection the units have. This is the principle of SE. According to the principle, we give the extending rule in this method as follow.

Extending Rule: A graph G , and G' is a subgraph of G . For any unit u , and $u \subseteq G-G'$. If $u \cap G' \neq \Phi$, u is the candidate unit for extending.

The process of the extending is iterative. The extending rule chooses the candidate unit for extending. In general, the candidate units are not unique. However, if all the candidate units are extended to the subgraph, the scale of the subgraph grows quickly. And the subgraph is

closed to the graph after several iterations. Thus, we needs to choose the unit for extending from the candidate set. Every time a subgraph only extends several units. The candidate units, which have more overlaps with the subgraph, are considered first. Choosing rule is described as follows.

Choosing Rule 1: A set S is the candidate units set, and the subgraph is denoted as G' . Choosing the units in the set S' and $S' = \{u \mid \max\{|u \cap G'|\}, u \subseteq S\}$.

The operation $|x|$ is also getting the number of the element in a set. The element of the set is the node. A unit is also a subgraph of the graph.

There could be more than one candidate unit with same most overlap. Therefore, we use a set to represent the choosing units. The number of units in the choosing set may be zero, one or more than one. The number shows the units which will be added in one time. If the number is zero, this is one case to stop the extending. If the result subgraph is satisfied for the need, the iterative also stops.

3.2 CONNECT-PATH DETECTION

In previous section, we described the principles and rules of SE. We also introduce the basic method of SE. There are several different implementations for different problems. In this section, we will detect connect-path between two aim-nodes in a graph. We conclude two implementations. One is a one-side extension. It starts extending from one node and stops when the other node is in subgraph. The other implementation is a two-side extension. It starts extending from both nodes and stops when the two extending subgraphs have overlap.

The former implementation chooses the units, which are strongest connection with subgraph to extend every time. This is easy to implement, but is blindly. If the other note is in the unit, which weaker connects with the subgraph, the first extending may omit that unit. This unit may be extended after many iterative times. The connect path between the two notes in the subgraph is longer. And sometimes, this method may miss that unit. The goal of latter implementation is clear. Every extension must consider the other subgraph. Moreover, both sides can extend in parallel. This reduces the running time of the extending. Here, we choose the latter extension method.

The two-side extension starts from two nodes. It extends the two subgraphs from two nodes. The initialize subgraph is the unit, which contains the aim-node. Figure 6 shows the initialize subgraphs of toy example in Figure 4.

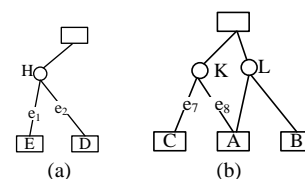


FIGURE 6 Initialize subgraphs of toy example. (a) shows the initialize subgraph of node e, and (b) shows the initialize subgraph of node a.

According to the analysis in Section 2.2, RDF-Bipartite graph is special. The connect-path must be in the same or higher layer. The candidate units must be in the same layer.

In each iterative, the extending processes from two nodes are not independent. They connect with each other. We do not only consider the connection between the candidate units and the subgraph, but also consider the subgraph extending from the other side. If the unit in both candidate sets of subgraphs extending from different sides, the unit connects both sides of subgraphs, it should be considered to be extended first. Then the subgraph contains two aim-nodes are found. The extending process could stop. According to the above analysis, we conclude another choosing rule as follows.

The graph is G , a subgraph from one side is G_1' , the other side is G_2' , the sets S_1, S_2 are the candidate units sets of G_1' and G_2' , respectively.

Choosing Rule 2: Choosing the units in a set S' and $S'' = \{u | \max \{|u \cap G_1'|, |u \cap G_2'|\}, u \subseteq S_1 \cap S_2, \text{ and } S_1 \cap S_2 \neq \Phi\}$.

Rule 2 is a supplement of rule 1. However, the two choosing rules do not have continuous relationship. When the candidate sets of two sides have some common units, the choosing rule 2 is available. And the final subgraph is $uUG_1'UG_2'$. The extending process stops. In other cases, the process extends and continue iterates following the rule 1.

After finding the subgraph, which contains two aim-nodes, the connect-path can be detected easily in the subgraph.

Let us continue to discuss the toy example in Figure 4. Figure 6 gives the initialized subgraphs for the toy example. Figure 7 shows the two candidate units sets of the two sides respectively.

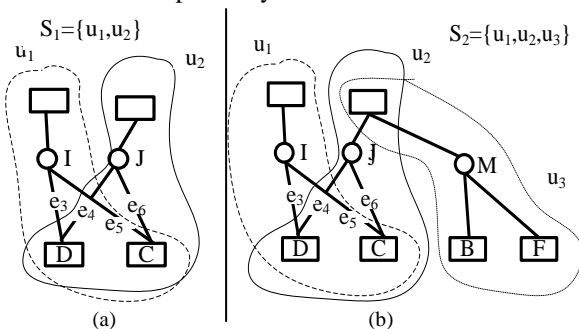


FIGURE 7 Candidate units sets. (a) is the candidate units sets of Figure 6(a). (b) is the candidate units sets of Figure 6(b).

In Figure 7, we can find that unit u_1 and unit u_2 belong to both candidate units sets. Therefore, u_1 and u_2 are both in the choosing set. Because u_1 has stronger connection with Figure 6(a), the u_1 is extended. The final subgraph is shown in Figure 8.

From Figure 8, we can find the connect-path between E and A is $(e_1, e_2, e_3, e_5, e_7, e_8)$.

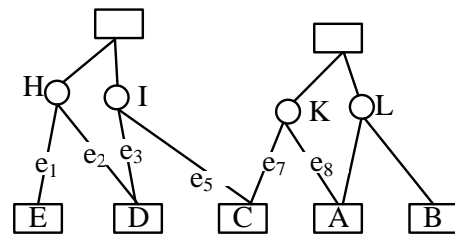


FIGURE 8 Final subgraph of the toy example

If the two nodes have weaker or no connection, the length of the connect-path between them should be larger or infinity, and the number of extending time is large or infinity. In addition, if the two nodes have weaker or no connection, they have smaller probability to be duplicates. Thus, we limit the length of the connect-path between two nodes. That is to limit the times of extending. When the number of extending times reaches the threshold value, we consider the two nodes distinct.

3.3 DUPLICATE DETECTION

However, if the two similar nodes have connect-path, they may be only similar and connective not duplicates. The probability of two similar nodes being duplicates is related to the connection between them. If two similar nodes have stronger connection, the probability that they are duplicates may be higher. Therefore, connect-weight of connect-path is also necessary in help judging. If we find the connect-path of two nodes, we can find all the units, which contain the nodes and edges of connect-path. We measure the connection between units instead of connection between nodes. The connective level between units is measured by connect-degree. The definition of connect-degree is given as follows.

Definition 3. Connect-degree. For any unit u_1 , and unit u_2 , the connect-degree is the number of nodes both in u_1 and u_2 , denoted as $d_c(u_1, u_2)$. That is $d_c(u_1, u_2) = |u_1 \cap u_2|$.

In RDF-bipartite graph model, each unit has four elements. According to the definition of RDF-partite model, if the connect-degree is more than 2, the two nodes are the same. So the maximal connect-degree is 2. Take Fig 7(b) for example, according to the definite 3, we can get that $d_c(u_1, u_2) = 2$, $d_c(u_2, u_3) = 1$, and $d_c(u_1, u_3) = 0$.

Because SE extends from two sides of the nodes, the more edges the connect-path has, the weaker connection the two nodes have. The connect-weight does not increase with the length of the connect-path. In addition, it may decrease while the length of the connect-path increases and the speed of decrease may increase with the length's increasing. It is hard to calculate connect-weight formally. Thus, we calculate the connect-resistance instead of connect-weight. The connect-resistance is the resistance of a path. Connect-resistance is in contrary to connect-weight. Connect-resistance is small, when the connect-weight is high. The connect-

resistance, denoted as r_c , between notes u and v is calculated by formula (1).

$$r_c(u, v) = 1 * [d_c(u, u_1) + d_c(v, v_1)]^{-1} + 2 * [d_c(u_1, u_2) + d_c(v_1, v_2)]^{-1} + \dots + n * [d_c(u_{n-1}, u_n) + d_c(v_{n-1}, v_n)]^{-1} \tag{1}$$

where $u_1, \dots, u_n, v_1, \dots, v_n$ are the nodes in the connect-path between u and v . And $u_n = v_n$. So $d_c(u_i, u_j)$ and $d_c(v_i, v_j)$ do not equal 0 simultaneously. And n is the number of extending times. It is noteworthy that the extending starts from two sides. So the length of connect-path is $(2n-1)$.

Because $d_c(u_i, u_j) = \{1, 2\}$, we can get $\min(r_c) = (1+2+\dots+n) * 1 / [2+2] = (1+n) * n / 8$; $\max(r_c) = (1+2+\dots+n) * 1 / [1+0] = (1+n) * n / 2$. The value area of r_c is shown in Figure 9.

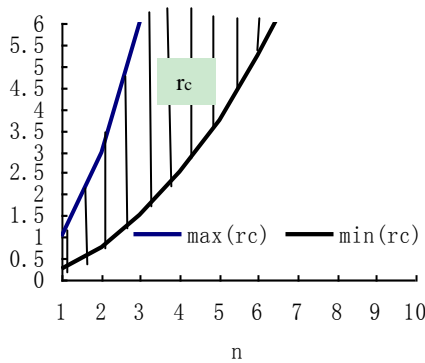


FIGURE 9 Value area of r_c

From Figure 9, the bigger n is, the longer the connect-path is, and then the connect-resistance must be bigger. And the slope of the connect-resistance's curve increases with the increasing of extending times. Along with the increase of the length of the connect-path, the connect-resistance could increase faster. So the trend of r_c conforms to reality.

If there is no path connecting two nodes, the connect-resistance is infinity. In Figure 8, the connect-path between E and A is $(e_1, e_2, e_3, e_5, e_7, e_8)$. In Figure 10, the units u, u_1 , and v contain all the nodes and edges in connect-path, and $d_c(u, u_1) = 2, d_c(u_1, v) = 1$.

Base on formula (1), the connect-resistance between E and A is $r_c(u, v) = 1 * 1 / (2+1) = 1/3$.

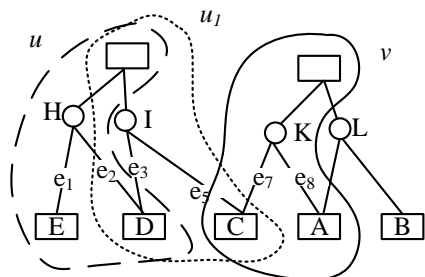


FIGURE 10 Connect units of Figure 8

If the connect-resistance is smaller than the threshold, the two nodes are considered as duplicates.

4 Experiment

In this section, we do experiments to examine the efficiency and accuracy of the proposed approach. All experiments are performed on an IBM eServer with a 1.25GHz Power4 processor and 4GB of memory, running Suse Linux Enterprise 10.0. All approaches are implemented and tested in Java.

The process of the proposed approach can be divided into three steps. The first step transfers the RDF data into RDF-Bipartite graph, like transfer fig 1. to fig 3. And then, calculate the similarities of the data pair wisely. The method of similarity calculation has been studied in [2]. All the similarity calculation pairs are in the same layer. And here, we still use the LFDW [2] to calculate similarity. If the similarity of the pair is higher than the threshold, the pair is the candidate pair. Then the candidate pairs need to be verified. The last step is to detect the connection between the data pair wisely. The connection between data is measured by the connect-path. If the connect-resistance of the connect-path between two data is smaller than a threshold, the two data are duplicates probably.

We use three popular metrics, recall, precision, and f-measure, to measure the efficiency of the approaches. In this paper, recall measures whether the approach detect all duplicates or not. Precision measures whether the detected duplicates is real duplicates or not. And F-measure is the harmonic mean of the precision and the recall rate. This metric measures the comprehensive performance.

The number of extending times is difficult to evaluate. If the number is too big, the connect-path is longer, and the connect-resistance is higher than the threshold. The two nodes are not duplicates. It wastes the running time. By conversely, if the number is too small, there are small pairs can find connect-paths. That may miss some duplicates pairs. From the analysis in Section III.C, the value of connect-resistance is in $[(1+n) * n / 8, (1+n) * n / 2]$. That means the connect-resistance is related with extending times n . The threshold thr is also related with the extending times. We can get the value of extending times satisfy with the in equation [2].

$$n \leq \left\lfloor \frac{\sqrt{32 * thr + 1} - 1}{2} \right\rfloor \tag{2}$$

If the threshold is known at the beginning, the number of extending times n is $\left\lfloor \frac{\sqrt{32 * thr + 1} - 1}{2} \right\rfloor$. And

if the threshold is unknown at the beginning, the value of the threshold is more difficult to be assigned. From Figure 9, the value of threshold is in a large scope, from

0.25 to infinity. If n is fixed, the threshold is fixed in the bound of $[(1+n)*n/8, (1+n)*n/2]$. Thus we assign $thr = [(1+n)*n/8 + (1+n)*n/2] / 2 = 5*(1+n)*n/16$.

Then we do experiments with different extending times. The results are shown in Figure 11 and 12.

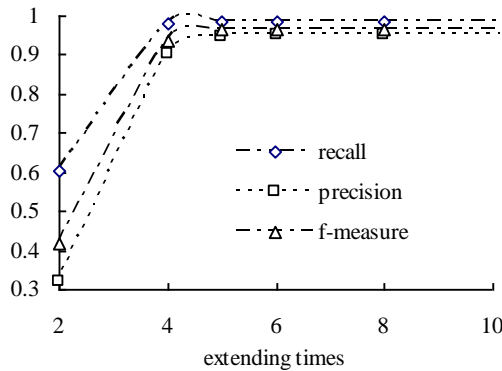


FIGURE 11 Detection results of different extending times

From Figure 11, we can find the f-measure is under 0.9 when n is small than 3. However, the f-measure increases quickly when n increases from 2 to 5. And after 5, the speed of increase is slow. The highest f-measure is get when n is 8. And when n is bigger than 8, the f-measure decreases a little, and keep the same value.

The f-measure is low in the beginning because n is too small to find the connect-path between duplicates. And along with the increase of the n , more connect-paths are found, so the f-measure increases quickly. Most duplicates are detected at this stage. When n continuously increases, there are only a few duplicates detected. Therefore, the increase speed is low. In addition, when all duplicate is detected, the f-measure is not change with n .

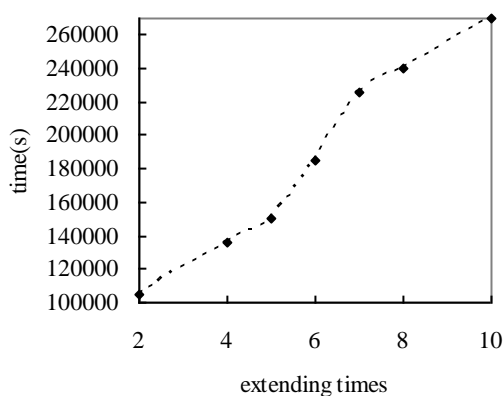


FIGURE 12 Costs of time in different extending times

Fig 12 shows the costs of time in different extending times. From figure 12, we can find the cost of time increase along the extending times increasing. This result is because more time is needed for subgraph extending to find connect-paths. And we can find the speed of increasing is not same. When n is smaller than 5, the

increasing is slowly. When n is larger than 5, the increasing becomes fast. These changes are happened by the reasons as follow.

According to the analysis of Figure 11, when extending times is 5, most duplicates are detected. And along with the extending times increasing, more pairs are found having connect-paths. From Figure 11, most duplicates are found connect-path in 5 extending. When the number of extending times is larger than 5, the detecting pairs most are not duplicates, while the subgraph becomes more complex and huge; more time is needed on extending and analysing for the huge subgraph.

From Figure 11, the highest f-measure is gotten when n is 8. However, the f-measure increase slowly when n is bigger than 5. And according to Figure 12, when n is bigger than 5, the running time increases quickly. So in conclusion, the option value of n is 5 in this RDF data.

We compare the proposed approach with the just similarity method. The proposed approach combines both similarity and connection. The result is shown in figure 13. In figure 13, for simple, we use 'S' to refer 'Similarity', and use 'C' to refer to 'Connection'. The note 'S+C' denotes the proposed approach.

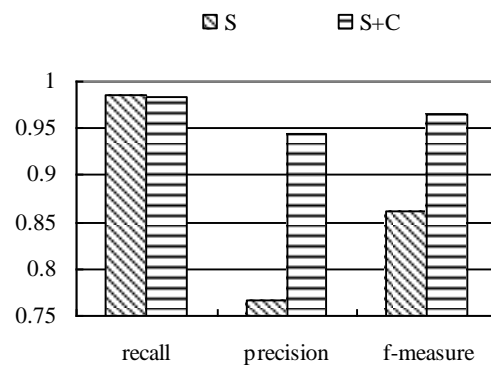


FIGURE 13 Result of comparison between S and S+C

From figure 13, we can find that, the precision of the proposed approach is higher than the method just considering similarity. This result shows that the connection among data helps for the duplicate detection. And the proposed method SE is effective for finding connection between data. In figure 13, we also find the recall of proposed method decreases tinely. This is because, the connection verifies after the filtering of similarity. So the recall of the proposed method is the same or smaller. We can assign the threshold of similarity smaller to get high recall, although more time costs for connection verifying.

5 Related work

Our research studies on solving the duplicate detecting in RDF data. This research relates with two main studies. One is duplicate detection, and the other is the model of RDF data.

Duplicate is the main inducement of data dirty. Duplicate detection is one kind of data cleansing. It is an important study in data mining. The basic method to detect duplicates is to compare entities. The main issue of compare entities is field matching [3], which could be achieved by recursive field matching algorithm [4], Smith-Waterman algorithm and R-S-W algorithm [5] etc. The above methods all consider the records themselves and omit the relationships among them. Recently, researchers shifted their attention to the associations among entities. Han et al. proposed an unsupervised learning approach using K-way spectral clustering that disambiguates authors in citations. It utilizes three kinds of citation attributes: co-author names, paper titles, and publication venue titles [6]. A general object distinction methodology is introduced in [7]. The approach combined two complementary measures for relational similarity: set resemblance of neighbour tuples and random walk probability, and then analysed subtle linkages effectively. Han et al. investigated two supervised learning approaches to disambiguate authors in the citations [8]. In [9], Kalashnikov proposed a domain-independent method. This method analysed not only object features but also inter-object relationships to improve the disambiguation quality. They used the shortest path algorithm to find the connect path connects two entities. And the path is measured by the association strength between the two entities. However, it is difficult to set the association strength between two entities correctly.

RDF is the W3C standard model for describing metadata. The RDF data also has the problem of duplicates. RDF data do not only represent the value of the data, but also represent the relationships among the data. Thus some studies represent the RDF data as graph. In [3], Klyne et.al. proposed a directed labelled graphs to represent the RDF data. A triple of RDF statement is a label edges. This model is easy to implement and represents the relationships among data clearly. However, if the relationships are complex, this model may miss some information. The RDF graph is different from a common graph. It is a hypergraph, because there may have more than one edge between two nodes. Therefore, in [1], Hayes proposed a undirected hypergraph model. In this model, each RDF statement is a hyperedge in the hypergraph. This model can represent the complex relationships among data. However, it cannot deal with the scale of RDF data, and is hard to more process on it. And in [1], they also proposed a bipartite graph model. This model transfers the hypergraph to a common bipartite graph. It is easy to do more processes. Our model is improved from this model. All the models introduced are using for semantic retrieval, like similar query and related query. They do not force on the duplicate detection.

6 Conclusion and future works

Nowadays, more and more resources are stored as RDF data. However, seldom researchers study on RDF data cleansing, although the problem of duplicates still exists in RDF data. According to that, in this paper, we proposed an approach to detect the duplicates in RDF data.

The proposed approach combines both similarity and connection among RDF data to detect the duplicates. And considering the complex of the association among RDF data, we proposed a new model for RDF, called RDF-Bipartite graph. This model is improved from Bipartite Statement-Value Graphs [1]. And we also proposed a method to find connect-path between two nodes in the RDF-Bipartite graph, called Subgraph-Extending method. This method does not find the connect-path directly, and instead to finding the subgraph contains both nodes. And introduce the connect resistance to help picking up the stronger connection pairs. This method is easy and efficiently to find the connection between two nodes.

We experiment the proposed method on publication datasets, and compare it with the traditional method. The results show that the proposed method improves accuracy and efficiency in detecting duplicates obviously.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61100055), The major projects of the National Social Science Foundation of China(Grant Nos.11&ZD189), Natural Science Foundation of Hubei Province (Grant No.500104), the Hubei Province Key Laboratory of Intelligent information processing and real-time industrial systems (Wuhan University of Science and Technology, Grant Nos. zNSS2013B013).

References

- [1] Amadis A, Gutierrez C 2007 A Directed Hypergraph Model for RDF In *proceeding of: Proceedings of the KWEPSY 2007 Knowledge Web PhD Symposium 2007*, Innsbruck, Austria, June 6, 2007 pp. 47–61
- [2] Huang L, Jin H, Yuan P, Chu F 2008 Duplicate Records Cleansing with Length Filtering and Dynamic Weighting *International Conference on Semantics, Knowledge and Grid 2008*, Dec. 4-6, 2008 95-102
- [3] Minton S, Nanjo C 2011 A Heterogeneous Field Matching Method for Record Linkage In: *Proceedings of the Fifth IEEE International Conference on Data Mining* 314-321
- [4] Atsuko Y, Yasunori Y, Kim In-Dong 2012 *Discriminative application of string similarity methods to chemical and non-chemical names for biomedical abbreviation clustering BMC Genomics* 13(3) S8
- [5] Bunke H 2012 On the Weighted Mean of a Pair of Strings *Pattern Analysis & Applications* 5 23-30
- [6] Anderson A 2010 Effective self-training author name disambiguation in scholarly digital libraries *Proceedings of the 10th annual joint conference on Digital libraries* 39-48

- [7] Yin X, Han J, Hu P 2010 Object Distinction: Distinguishing Entities with Identical Names In *ICDE 2010: IEEE 23rd International Conference* 1242-1246
- [8] Han H, Giles L, Zha H 2012 Two Supervised Learning Approaches for Name Disambiguation in Author Citations *JCDL '12* 269-305
- [9] Kalashnikov D, Mehrotra S 2012 Domain-independent data cleaning via analysis of entity-relationship graph *ACM Transaction on Database Systems* 31(2) 716-767
- [10] Klyne G, Carroll J 2004 *Resource description framework (RDF): Concepts and Abstract Syntax* W3C Recommendation <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/> 10 February

Authors



Huang Li, born in 1982, Wuhan, China

Current position, grades: lecturer of computer science

University studies: PhD in computer science

Scientific interest: Data management, Semantic web and knowledge.

Publications: 10

Experience: Li Huang is a PhD in computer science. Currently, she is a lecturer of computer science of the School of Computer Science and Technology, WUST, Wuhan, China. Her research interests include data management, semantic web and knowledge.

A dynamic MAC protocol for wireless sensor network based network traffic monitoring and feedback

Yang Xianhui, Ren Honge^{*}, Jing Weipeng

The College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China

Received 1 January 2014, www.tsi.lv

Abstract

WSN application in coal mine gas monitoring protects safety production. In order to save energy consumption, SMAC use periodic listen and sleep mechanism, against the problem of fixed duty cycle causes increased delay, this paper puts forward a new MAC protocol—DSMAC(Dynamic SMAC)which can feedback the channel's congestion level reasonably and adjust the duty cycle dynamically. The dominant idea of the new scheme is to record the packet amounts in the sending stage by sensor nodes, and according to the packet sending rate every sensor node can change its listening time adaptively when synchronization cycle finished to minimize delay and collisions, saving energy and channel resources. The result of simulation shows that compared with other MAC protocols, the DSMAC protocol can improve network throughput and reduce energy consumption in coal mine gas monitoring environment.

Keywords: WSN, periodic listen and sleep, duty cycle, DSMAC

1 Introduction

With the rapid development of radio systems, on-chip system, wireless communications and low-power embedded technology, wireless sensor networks have emerged, and its characteristics of low power, low cost, distributed version and self-organization brings a revolution in information awareness [1-3]. Therefore, the applications of wireless sensor network in forest resources monitoring system will greatly improve the accuracy of monitoring data. Real-time acquisition of environmental information by sensor nodes, and being processed by the embedded systems, then transmitting the information to the monitoring terminal outside the wells through self-organizing wireless networks in multi-hop relay way, that can effectively compensate for the shortcomings of wired devices, and has many advantages such as cheap, portable, high reliability, scalability, etc. [4].

Wireless sensor nodes are generally battery powered; when the energy is exhausted and cannot replace the power supply, one of the major technical challenges ,the wireless sensor network facing, is that nodes should meet the requirements of low power consumption, making access to the entire network to work as long as possible [5].Media Access Control (MAC) protocol determines the use of radio channel to allocate the limited resources of wireless communication between the sensor nodes, then how to design energy-efficient, low-overhead, high but also scalability fair MAC protocol has become an urgent research topic in the field of wireless sensor networks currently.

Currently, wireless sensor networks with different characteristics and specific applications of MAC protocols have been proposed [6-10]. BMAC [6] uses the extended protocol preamble and low power listening technique to achieve low power communications, channel ruling by the clear channel assessment techniques, and the BMAC does not need to share scheduling information. Therefore, it has a higher throughput and lower latency However, not much advantage gained in reducing the energy consumption. X-MAC [7] protocols shorten the length of the leader sequence, while introducing a handshake mechanism to further reduce the energy consumption of transmitting the preamble sequence. The experiments show that X-MAC exhibit excellent energy efficiency, throughput, and delay in different network conditions, but its computational methods are very complex.

In order to meet the need of low-power sensor networks, Wei Ye et al proposed SMAC [11] protocol having a sleep mechanism. Based on the analysis of the working process of SMAC protocol, against the problem of fixed sleeping time, this paper puts forward a new MAC protocol—DSMAC (Dynamic SMAC) which can feedback the channel's congestion level reasonably and adjust the sleeping time dynamically.

The improved protocol has increased the throughput in coal monitoring environment, reduced the rate of package loss and energy consumption, showing good network performance.

^{*}Corresponding author - E-mail: 14108791@qq.com

2 Analysis of SMAC protocol

Based on IEEE802.11 [12] protocol, SMAC is a competition-based distribution MAC protocol to meet the need of low-power sensor networks. It can make the sensor find the neighbouring nodes without the scheduling of the master node, and reasonably arrange the time occupied by channel, having good scalability.

In order to reduce the energy consumption by nodes, SMAC using periodic listener / sleep mode of operation, turn off the wireless device node and set the timer to wake up and listen to other nodes during sleep. The advantage is the fixed sleeping time of each node, so it can't be well adapted to the changes of network traffic, resulting in unnecessary energy consumption. The time of a periodic listener / sleep is one frame, during which presents different working conditions, and the listening period is divided into three parts shown in Figure 1.

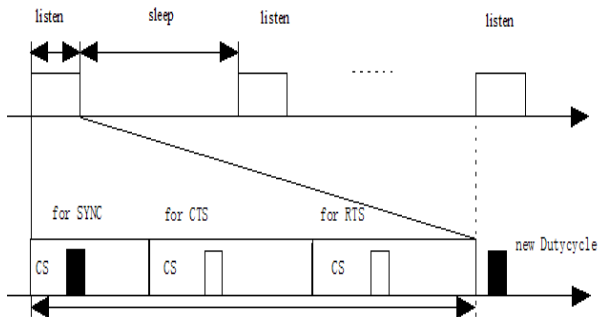


FIGURE 1 SMAC periodic listener / sleep and SYNC cycle

Compared with IEEE 802.11 protocol, SMAC greatly reduces the energy of nodes, thereby increasing the life of wireless sensor networks. The amount of saving energy consumption can be measured by duty circle, defined as follows:

$$Duty = \frac{T_{listen}}{T_{frame}} \tag{1}$$

Thereby the saving energy ratio in SMAC protocol is:

$$E_{smc} = \frac{T_{sleep}}{T_{frame}} = 1 - \frac{T_{listen}}{T_{frame}} = 1 - Duty \tag{2}$$

$$T_{listen} = T_{SYNC} + T_{RTS} + T_{CTS} \tag{3}$$

$$T_{frame} = T_{listen} + T_{sleep} \tag{4}$$

It can be seen that the duty cycle marks the size of the node energy consumption. The larger the duty cycle is,

the longer the node works, the greater the energy consumption is; the smaller the duty cycle is, the longer the node sleeps, though effectively reducing the energy consumption caused by the listener, the delaying time caused by the sleep will be longer. Below is the analysis on the issue.

For SMAC protocol the transmission of data packets from the source node to the next node will be affected by the delay, the paper analyses the delay of the packet when the network load is extremely light. Assuming that only one packet is transmitted in the network, it will not produce back off window delay and queuing delay. Moreover, the propagation delay and processing delay is very small, so it can be negligible. Therefore, the packet delays $D(k)$ in the k ($k=1,2,\dots,N$) hop only need to consider sleep latency ($T_{s,k}$), competition delay ($T_{c,k}$) and transmission delay (T_{tr}).

The time simultaneous packet transmitted T_{SYNC} , so the packet delay $D(k)$ in the k ($k=1,2,\dots,N$) hop is:

$$D(k) = T_{s,k} + T_{SYNC} + T_{c,k} + T_{tr} \tag{5}$$

Then can be obtained:

$$D(k) = T_{frame} - T_{c,k-1} + T_{c,k} \tag{6}$$

The total delay after N-hops:

$$D(N) = \sum_{k=1}^N D(k) = D(1) + \sum_{k=2}^N D(k) = D(1) + \sum_{k=2}^N (T_{frame} - T_{c,k-1} + T_{c,k}) \tag{7}$$

After finishing it:

$$D(N) = T_{s,1} + T_{c,N} + T_{tr} + (N-1)T_{frame} \tag{8}$$

Since the source node generates a packet at random, $T_{s,1}$ is evenly distributed within $(0, T_f)$, thereby the average delay of SMAC:

$$E[D(N)]_s = NT_{frame} - \frac{1}{2}T_{frame} + T_{tr} + T_c \tag{9}$$

Without introducing sleeping mechanism, the average delay of IEEE802.11:

$$E[D(N)] = T_{SYNC} + T_c + T_{tr} \tag{10}$$

T_{SYNC} can be negligible with a short frame. In SMAC protocol T_{listen} is fixed, so in order to reduce more energy consumption, we should use longer T_{frame} to reduce $Duty$. However, the delay was linearly increased, which is the contradiction between the delay and energy consumption.

Meanwhile SMAC protocol uses a fixed duty cycle, when the networks load changes, adaptive capacity is relatively poor, and the delay is a serious problem. Therefore, how to dynamically adjust the duty cycle is an important direction sensor network MAC protocol design. So that it can adapt to changes in network load, which can reduce latency and increase throughput as well as maintain efficient energy utilization.

3 Relevant improvements about the MAC protocol

3.1 T-MAC PROTOCOL

T-MAC [13] (Timeout-MAC) protocol is proposed which is based on SMAC protocol. T-MAC protocol based on the unchanged cycle length, according to traffic to adjust activity time dynamically, T-MAC protocol reduces the idle listening time by means of sending a message with the burst mode. T-MAC uses the same frame size with SMAC and nodes are periodically awaked to listen. If there is no activation events that occur in a given TA (Time Active, each cycle determines the minimum idle listening time), the activity will end.

Based on traffic to adopt the dynamically changeable duty cycle, T-MAC protocol saves energy consumption, extending the life span of the network compared to SMAC protocol. However, because the node, which is beyond the scope of monitoring and needs to participate in data forwarding, cannot get timely information to extend the activation period and may be transferred to sleep early, resulting in "early to bed" problem. Thus, this reduces network throughput and increases network latency. In order to resolve the "early to bed" issue, T-MAC protocol raised two programs .One is request-to-send in future and the other is full-buffer priority, but neither is ideal.

3.2 AC-MAC PROTOCOL

AC-MAC is a flow-based adaptive SMAC protocol. It is the number of packets on the node of MAC layer that is the criterion to evaluate the node load. Moreover, based on the node load, it is to adjust the amount of SMAC within a communication cycle in order to adjust the duty cycle. It is shown in Figure 2.

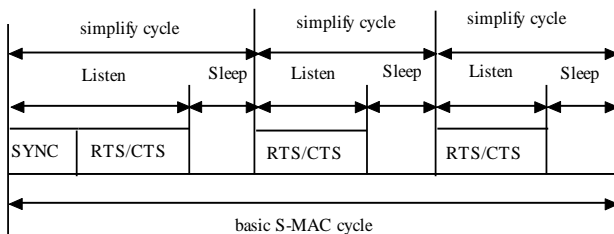


FIGURE 2 The AC-MAC new sleep cycle

The basic cycle is classified as several new simplified cycles on the basis of node queue load. Large loaded nodes can be awaked many times to get more

opportunities to communicate without having to wait until the next time slot. Nevertheless, added to new simplified cycles, nodes can be frequently status switching and consumptive energy cannot be ignored.

Assume that node hibernation for energy is E_{sleep} , activity for energy is E_{listen} , the conversion time for the listen to the sleep is T_{ls} , the conversion time for the sleep to the listen is T_{sl} , then consumption of using a periodic listener /sleep strategy consumption $E_{listen / sleep}$ is:

$$E_{listen / sleep} = \frac{1}{2}(E_{listen} + E_{sleep})(T_{ls} + T_{sl}) + E_{sleep}(T_3 - T_{ls}). \quad (11)$$

The state of transition and energy is shown in Figure 3

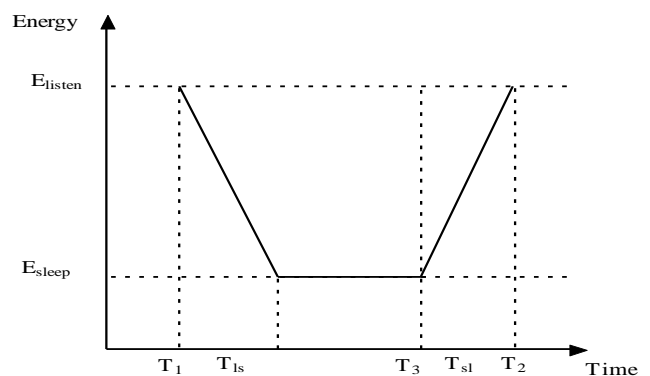


FIGURE 3 The state of transition and energy

3.3 MAC-ADCC PROTOCOL

ADCC (Adaptive Duty-cycle Based Congestion Control) By means of the dual monitoring the residual energy and network load conditions and feedback, it can adjust the duty cycle adaptively and reduce the collision and data retransmission. The results show that, ADCC-based MAC protocol reduces the average network latency in high-speed network environment and improves network throughput.

Meanwhile the literature, it is respectively proposed LQ-Tracking algorithm and TA-MAC to adjust the duty cycle dynamically to reduce collisions and to lower latency [16, 17]. However, because the introduction of more parameters and adjustment mechanism are quite much, so that the complexity of the algorithm increases certain energy consumption, and still needs further improvement.

3.4 DSMAC PROTOCOL

Based on the analysis of the above algorithm of duty cycle of several MAC protocols, we propose DSMAC. It is a SMAC-based protocol. Moreover, based on the rate of node synchronization cycle to send data packets, it can forecast network load in the future periods and adjust duty cycle dynamically. DSMAC solve in the dynamic changes in network load, SMAC protocol leads to the

listening time adjustment ineffectively because of the fixed duty cycle mechanisms, resulting in the problem of excessive delays. However, DSMAC resolves this problem. And according to theoretical analysis, DSMAC does not cause excess energy consumption, and the algorithm is easy to implement and has a good network performance.

Here concepts of synchronization cycle and the data packet transmission rate and adjustment rules of duty cycle are defined and analysed.

Definitions 1. Adjacent nodes to regularly update listeners / sleep scheduling information cycle by sending SYNC frame is synchronized. With expression of Synchronization the default value is 10.

Synchronization cycle length is T_{SCYCLE} , the unit is s , the specific calculation method is as follows:

$$\begin{aligned} T_{SCYCLE} &= SYNCPERIOD \times (T_{listen} + T_{sleep}) = \\ &= SYNCPERIOD \times T_{frame} \end{aligned} \quad (12)$$

Definition 2. The number of transmission packets per unit time is the packet transmission rate, by means of ∂ . It is defined as:

$$\partial = \frac{totalpackets \times 50 \times 8}{T_{SCYCLE} \times 1000} \quad (13)$$

The unit is $kbit/s$. the number of transmission packets in the synchronizing cycle is showed by $totalpackets$, the size of a data packet is 50 bytes.

Definition 3. Duty cycle adjustment rules.

In DSMAC, follow the cycle length of SMAC, about 1.4 s, bandwidth is 20kbps.

Therefore, the transmission time required T_{bit} for a byte is

$$T_{bit} = \frac{1}{20 \times 10^3} \times 8 = 0.4ms \quad (14)$$

In NS-2, SMAC protocol was specified in the packet up to 1000 bytes. In order to facilitate the calculation, it requires all packet sizes are 250 bytes, letting time of each packet is approximately $250 \times 0.4ms = 0.1s$.

When ∂ equals to 1, s , the number of packet data being sent every second equals to $1000 / (250 \times 8) = 0.5$. Then the number of packet data being sent during every period of T_{frame} is 0.7. Thus, the time is 0.07s and the duty ratio $Dutycycle$ is

$$Dutycycle = \frac{T_{listen}}{T_{frame}} \times 100\% = \frac{0.07}{1.4} \times 100\% = 5\% \quad (15)$$

When ∂ equals to 2, the number of packet data being sent every second equals to 1. Then the number of packet

data being sent during every period of T_{frame} is 1.4. Thus, the time is 0.14s and the duty ratio $Dutycycle$ is 10%.

From the above, it can be concluded that the duty ratio and packet sending rate are in linear relationship, that is

$$DUTY_{new} = a\partial + b \quad (16)$$

Therefore, $a=5$. As conflicts are not included in the calculation, the adaptive duty ratio parameters of adjusting mechanism b is also set to be 5 considering circumstances of conflicts. Then, the following magnitude comes:

$$DUTY_{new} = 5\partial + 5 \quad (17)$$

Set the synchronous timerds $DSMACGeneTime()$ after starting the wireless sensor network node wireless sensor network node wireless sensor network node wireless sensor network node, the cycle is T_{SCYCLE} . During every time slot, the timer minus 1, until the timer becomes 0. Then restart timer for the next monitor. During the process, the nodes should be recorded as the number of packet data being sent $totalpackets$ in synchronous period.

And work out the packet sending rate ∂ according to the formulas of (12) and (13). After that, adjust the duty ratio to update synchronously in the timer $DSMACCGeneTimer()$ trigger event $DSMAC :: setMySched()$ function according to formula (16), in the meanwhile, reset the number of packet data to 0.

The specific implementations are as follows:

Algorithm DSMAC

```

1: start DSMACGeneTimer()
2: {
3: while(mhGene_.sched(TSCYCLE))
4: {
5:   if(sendpackets)
6:     totalpackets++;
7: }
8: calculate packet sending rate ∂;
9: DSMAC :: setMySched();
10: {
11:   refresh_duty();
12:   {
13:     recalculate DUTYnew;
14:     DUTYnew = 5∂ + 5;
15:   }
16:   update Dutycycle;
17:   totalpackets = 0;
18: }
19: }

```

```

Algorithm mhGene_.sched()
1: start mhGene_.sched
2: {
3: while( $T_{SCYCLE}$ )
4: {
5:    $T_{SCYCLE}--$ ;
6: }
7: }
8: DSMAC :: setMySched();
9: }
    
```

Synchronous cycle period being taken as the cycle of timer, it is to prevent that the remaining times are not synchronized among the nodes within the current luster in the synchronous cycle after the node is changing the duty ratio, which leads to the problem in node communication.

Based on previous packet sending rate δ of nodes in the synchronous cycle, DSMAC can predict the condition of the load of nodes in the following periods, which contributes to the adjustment of the duty ratio. If the present packet sending rate is high, increasing the duty ratio in linear according to the principle of adjusting the duty ratio in order to solve problems. The problems are as follows: delay increases and throughput decreases, which results from source node's waiting for the destination node waking up from sleep. If the present packet sending rate is low, decrease the duty ratio in linear to decline the interception time, save the energy and raise the utilization of channel and energy efficiency.

In DSMAC protocol, the length of node cycle T_{frame} is fixed, and can adaptively change the interception time T_{listen} according to network load so as to make dynamic adjusting in the duty ratio *Dutycycle*. It solves the problem of contradiction between delay and energy consumption in SMAC protocol.

The interception time of DSMAC protocol needs no active event to adjust, so there is no concern about the problem of early to bed led by T-MAC protocol.

DSMAC has only one state transition in a cycle, solving the problem of excess energy consumption caused by state transition of many times in AC-MAC. Besides, it is simple to accomplish and does not increase in algorithmic complexity. Therefore, DSMAC can achieve the aims of adapting to the change of network load, decrease the delay effectively, increase throughput capacity and use energy efficiently at the same time.

4 The simulation experiment and results analysis

In order to verify the validity of DSMAC, this paper make it, together with original SMAC protocol and IEEE802.11 protocol conduct the simulation experiment in the same network environment, and analyse the results of performance. The simulation tool is NS-2[18], developed by Berkeley University.

Arrange 101 wireless sensor nodes randomly in the scope of $1200 \times 1200m$, and number them from 0 to 100.

Form a pair of CBR data stream transmission links respectively between node 1 and 100, node 2 and 99. In order to know the influences towards network performance in different network flow, the starting time of the two CBR data streams differs. The starting time of the first data stream is 50s and the packet sending rate is 1kbps at the start. The packet sending rate increase 1 for every 200s until up to 5kbps. At 1050s, the two data streams starts together, and the packet sending rate increase from 6kbps to 10kbps. The total time of simulation is 2100s. Then, a scene where the load is growing is constituted to demonstrate the relevant performance of DS-MAC mechanism in network with load fluctuation. Parameter setting Parameter settings in parameter setting NS2 are shown in Table 1.

TABLE 1 Simulation Parameter

Simulation Parameter	Parameter Values	Simulation Parameter	Parameter Values
Energy Model	Energy Model	Idle Power	0.36w
Propagation Type	TwoRay Ground	RxPower	0.36w
Initial Energy	10000J	TxPower	0.65w
BandWidth	20kbps	SleepPower	5.0e-5w
Packet Size	250bytes	Transition Power	0.05w
Numbers of node	101	Transition Time	0.0005s
Initial dutycycle	10%	RXThresh_	3.65e-10s

The thesis demonstrates the network simulation to SMAC Protocol, DSMAC and IEEE802.11 Protocol. Next will be the contrastive analysis of network throughput, the efficiency of network, the average network time-delay and packet loss probability under the situation of the growing of load in the increasing of simulation time.

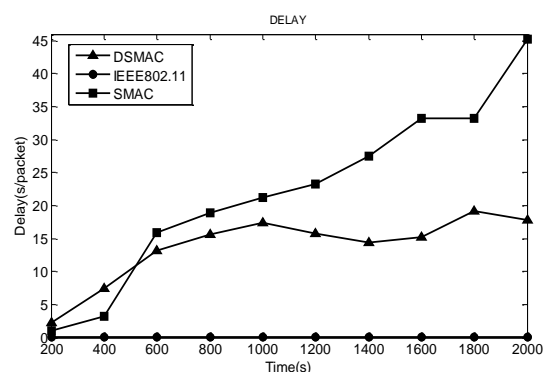


FIGURE 4 The comparison of Network Time-delay of Three MAC Protocols

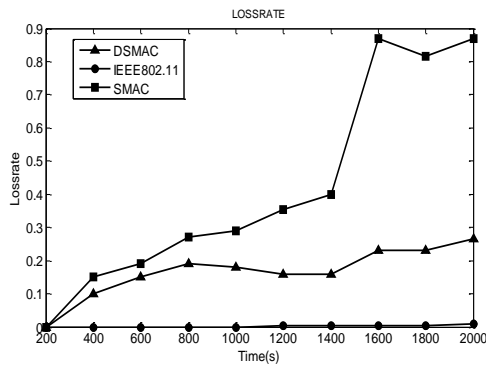


FIGURE 5 The comparison of Packet Loss Probability of Three MAC Protocols

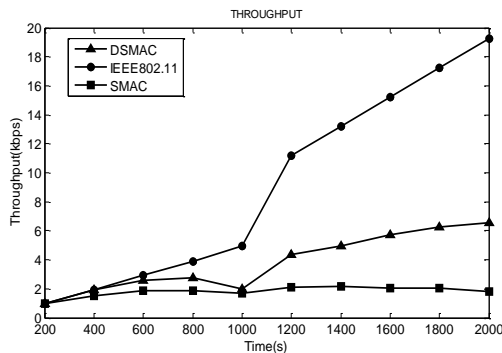


FIGURE 6 The Comparison of Throughput of Three MAC Protocols

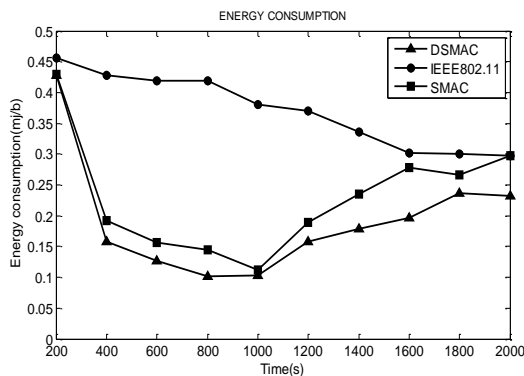


FIGURE 7 The Comparison of Energy Consumption of Three MAC Protocols

From Figure 4 and Figure 5, it can be seen that in the whole process of simulation, the average of time-delay and packet loss probability of IEEE802.11 keep in a low level (approximately to 0). Without the intercept mechanism of sleep cycle, the node is always in intercept condition. If data reception or transmission is needed, the wake-up of node is unnecessary, thus the average of time-delay is in minimum. The average time-delay and packet loss probability of SMAC Protocol are in remarkable increasing tendency, which results from that with the continuous activation of high package delivery, the efficiency of package delivery of the whole network is also increasing and more and more data package will complete transmission in 10% of fixed space, leading to the originating code waiting for the wake-up of

destination code and the fierce network competition in serious collision as well as the remarkable increase of the average time-delay and packet loss probability of the whole network. After the utilization of DS-MAC Protocol in which the regulatory mechanism is based on the network traffic control, because the node can accumulate the duty ratio according to state of the data package delivery. When the number of package is increasing, the node can increase its duty ratio as well as the intercept time, making the time-delay and packet loss probability of network highly improved.

The Figure 6 illustrates network throughput of the IEEE802.11DCF, which is always in intercept condition, is very large in unit time. Because the SMAC Protocol take the fixed Duty Ratio Mechanism, when the network flow becomes lager, many groups wait in line, the intercept cycle of node, which wants to transmit the group, cannot adjust itself roundly and the re-transmission will increase after collision, making the throughput stays at the relatively low level. Through the supervision of packet delivery, the DS-MAC Protocol can adjust the duty ration of node according to the change of network flow to improve the network throughput and secure the coal monitor work.

As can be seen from Figure 7, IEEE802.11, SMAC DSMAC bit and high energy consumption in the initial phase, which is the initial stage because only one data stream and a low transmission rate CBR, during which time most of the nodes in the no data transmission is in idle listening state. Around 1000s, SMAC bit consumption to a minimum, and then as the load increases, the node duty cycle is fixed, fierce competition caused large amounts of data packet retransmission, while the source node needs to wait for the destination node wakes up from sleep, lead to networks a tendency to increase the energy consumption. Use the dynamic adjustment of the duty cycle of DS-MAC protocol throughout the simulation stage, the bit consumption of energy is always lower than SMAC. DS-MAC protocol node contract rate increases with increasing duty cycle, although more time to be active, but at the same time significantly improve latency and throughput in the network environment, real-time monitoring of coal node protection work and effectiveness.

5 Conclusions

WSN will be introduced in monitoring forest resources to build a comprehensive, real-time environmental monitoring system. We made an in-depth research and improvement and proposed DS-MAC for the MAC layer technology of WSN to be more suitable to the forest resources monitoring system. DS-MAC protocol has positive significance for forestry internet of things, which is based on WSN.




Acknowledgment

The work described in this paper is supported by Supported by Natural Science Foundation of Heilongjiang Province of China (ZD201203/ C1603)

References

[1] Fengyuan Ren, Hai ning Huang, Chuang Lin 2003 *Journal of Software (in Chinese)* 14(7) 1282-91
 [2] Jianzhong Li, Jin bao Li 2003 *Journal of Software (in Chinese)* 14(10) 1717-27
 [3] Potte G, Kaise W 2000 *Communications of the ACM* 4(5) 551-8
 [4] Estrin D, Govindan R, Heidemann J 1999 *Proceeding of the 5th annual ACM/IEEE international conference on Mobile Computing and Networking* New York: ACM 263-70
 [5] De pei Qian, Shihan Li, Yi Liu et al 2009 *High-tech communications (in Chinese)* 19(5) 441-5
 [6] Polastre J, Hill J, Culler D 2004 *Proceeding of the 2nd ACM conference on Embedded Networked Sensor Systems (SenSys)* Baltimore 95-107
 [7] Buettner M, Yee G, Anderson E, Han R 2006 *Proceeding of the 4th ACM international conference on Embedded Sensor Systems(SenSys)* New York: ACM Press 307-20
 [8] Ceken C 2008 *Elsevier Science* 30(1-2) 20-31

[9] Pei H, Chen W, Li X, et Chen H 2010 *Proceeding of the 18th International Workshop on Quality of Service (IWQoS)* Beijing 1-9
 [10] Guoqiang Z, Yanhua G, Jiangtao F, Shengyu T 2010 *Proceeding of International Conference on Communications and Mobile Computing (CMC)*. Shenzhen 224-8
 [11] Wei Y, Heidemann J, Estrin D 2002 *Proceeding of the 21th annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)* New York USA 1567-76
 [12] *IEEE Computer Society LAN MAN Standards Committee IEEE Std 802.11-1999* 1999 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications
 [13] Dam T V, Langendoen K 2003 *Proceeding of the 1st International Conference on Embedded Networked Sensor Systems* Los Angeles, CA, USA 5-7
 [14] Jing A, Jingfei K, Damla T 2004 *Proceeding of the 9th IEEE Symp on computers and communications (ISCC)* Alexandria, Egypt 214-9
 [15] Dongho L, Kwangsue C 2010 *IEEE Transactions on Consumer Electronics* (56) 42-47
 [16] Vigorito C M, Ganesan D, Barto A G 2007 *Proceeding of 4th Annual IEEE Communications Society, Sensor, Mesh and Ad Hoc Communications and Networks* San Diego:CA 21-30
 [17] Hao C, Hao L 2008 *Proceeding of 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian* 1-4
 [18] Huaqi shang, Suili Feng, Lijiao Qing 2010 *NS network simulation and emulation protocol* Beijing:People Post Press

Authors	
	<p>Xianhui Yang, 05 05 1980 Heilongjiang, China</p> <p>University studies: Northeast Forestry University, Master Degree of Computer Application (2011), PhD studies Scientific interest: Wireless Sensor Networks.</p>
	<p>Honge Ren, 11 1962, Jilin, China</p> <p>Current position, grades: professor of The College of Information and Computer Engineering, Senior Member of the CCF University studies: Northeast Forestry University, PHD degree (2006) Scientific interests: research interests include different aspects of Artificial Intelligence and Distributed Systems</p>
	<p>Weipeng Jing, 01 1979, Heilongjiang, China</p> <p>Current position, grades: Associate Professor in Northeast Forestry University, a member of the CCF Scientific interests: His research interests include modelling and scheduling for distributed computing systems, system reliability estimation, fault tolerant computing and system reliability, distributed computing</p>

A concurrent MAC protocol based on location information in wireless sensor networks

Guoyan Yang, Xin Guan *

College of Information Science and Technology Heilongjiang University, Harbin P. R. China, 1500801

Received 1 January 2014, www.tsi.lv

Abstract

The traditional CSMA MAC node simply blocks its transmission if the medium is sensed to be busy. Thus, it is inefficient in terms of the network throughput due to overcautious estimation of the interference. In this paper, we propose a novel location-aware medium access protocol for data intensive wireless sensor networks. In this protocol, the contending nodes make use of their location information to achieve the concurrent transmission of exposed terminal so as to reduce collisions and improve the overall performance. We evaluate it in terms of delay and throughput and compare it with S-AMC using simulations. Results show that the proposed MAC protocol can take advantage of the location distribution of nodes to improve the average throughput of the network, reducing data transmission delay, and effectively improving the efficiency and performance in data intensive wireless sensor networks compared to S-AMC.

Keywords: exposed terminal, MAC protocol, wireless sensor networks, concurrent transmission.

1 Introduction

In wireless sensor networks, different users try to access the same medium. Thus, contention becomes a limiting factor of the MAC layer performance, and managing spatial reuse is a critical issue. IEEE 802.11 MAC protocol [1] is becoming the most popular protocol for wireless networks. As all carrier sense multiple access with collision avoidance (CSMA/CA) based MAC protocols, DCF suffers from hidden and exposed terminal problems. The basic mechanism of DCF is that a station may transmit only when the medium is sensed to be idle since any station may cause interference to an ongoing transmission occupying the medium. More specifically, if a station has data to transmit but a busy carrier has been detected, a blocking mechanism is performed (i.e., the transmission is delayed).

On the other hand, it was shown that the DCF is inefficient with respect to the network throughput [2]-[6]. This is because the unnecessary blocking of concurrent transmission is occurred when a station senses the medium busy or it receives a channel reservation frame from other station. Actually, in many cases, the station accessing the channel may not produce sufficient interference to inhibit ongoing transmission to the intended receiver.

To deal with this problem of the DCF algorithm, various intensive research works have been reported in the literature. Among them, we will review several previous works closely related to our study. First, the authors of [2] has adjusted the timing of RTS/CTS/DATA/ACK frame sequence and attempted to

synchronize one hop neighbours for concurrent transmission. The authors of [3] proposed a new blocking condition, which obstructs the transmission only when both RTS and CTS frames are received. However, these approaches have a main drawback that is feasible only for some special network topology. Secondly, [4] introduced the method that embeds the measured signal to interference and noise ratio (SINR) using the RTS frame into the CTS frame in order to exploit spatial resource. However, in that schemes, the only the DATA frame was considered during the concurrent transmission. In practice, all DATA frames in medium access control (MAC) layer get an acknowledgement (ACK) receipt to provide adequate link reliability. In addition, they did not consider the contention between opportunistic transmissions that causes the performance degradation. Finally, per-packet power control mechanism was suggested to leverage the channel spatial reuse (e.g., [5]). According to these methods, the concurrent transmission is possible in the same vicinity of a receiver by the local broadcast of collision avoidance information over a separate control channel. However, these schemes have a constraint that the station must be equipped with two transceivers and needs two orthogonal separate channels.

Location awareness of wireless nodes is increasingly common in many wireless network applications. In this paper, we propose a new approach to improve spatial reuse in multihop wireless sensor networks. This approach exploits location information to schedule concurrent transmissions. The objective is to allow an exposed node to schedule concurrent transmissions to improve network throughput and delay performance. We

*Corresponding author - E-mail: hdyanguoyan@gmail.com

assume that each node has its location information, either through the Global Positioning System or many available localization algorithms [6]-[7]. We further assume that each node is able to exchange location information with its neighbours.

2 Related work

IEEE 802.11 uses the so-called virtual carrier sensing mechanism to resolve the hidden-terminal problem. However, the so-called “exposed-terminal” problem still remains.

In Figure 1 we suppose node A is transmitting to node B and after some time node C wants to transmit to D. According to the CSMA protocol, node C senses the medium, finds that node A is transmitting and waits (node C) until node A is finished with its transmission. This occurrence is known as exposed terminal problem which is responsible for degrading the network performance [8], because from the above scenario we find that node C could transmit to node D without collision and hence save a significant amount of time.

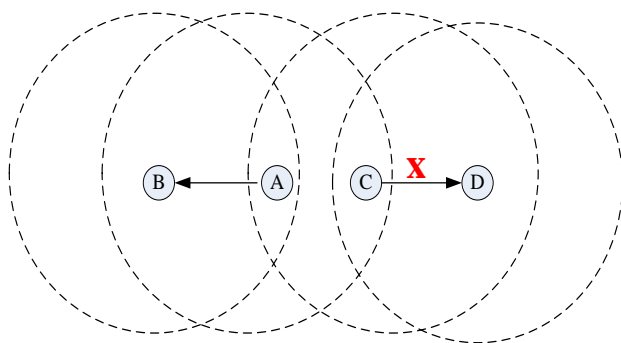


FIGURE 1 Exposed terminal problem

Recently there have been considerable efforts on improving the throughput of wireless networks by enhancing DCF [9-16]. A relatively simple scheme is presented in [17], which enables nodes to identify themselves as exposed terminals and opportunistically schedule transmission of a small frame without RTS-CTS exchange. In [13], Acharya et al. proposed a new MAC protocol, called MACA-P. The key idea of this protocol is to introduce an extra gap between the RTS/CTS frames and the subsequent DATA frames in addition to the short inter-frame space (SIFS) of the IEEE 802.11 MAC protocol. This extra gap allows all the neighbouring nodes to exchange the RTS/CTS frames for the purpose of concurrent transmission. After successfully exchanging the RTS/CTS frames by the end of this extra gap, the links with concurrent transmission opportunity, such as the links $B \rightarrow A$ and $C \rightarrow D$ in Fig. 1, can start transmitting their DATA frames. For concurrent transmissions, extra information bits are added in the RTS and CTS frames to indicate the start time of the DATA frame and the ACK frame. Hence, the two

concurrent transmission links can synchronize their starting time. Similarly, the virtual-carrier exposed node issue can be solved by the MACA-P MAC scheme.

The improvement of concurrent transmission opportunity from the MACP-P MAC protocol comes at the price of memory cost and incompatibility. The wireless ad hoc network adopting the MACA-P MAC protocol requires a larger memory size for storing the scheduled transmission time of all the neighbouring nodes. More importantly, the MACA-P MAC protocol is not compatible with the IEEE 802.11 DCF MAC protocol.

In [18] Shukla et al. proposed a parallel-MAC (P-MAC) protocol to increase the concurrent transmission opportunity of a short packet together with a long packet. The basic idea of the P-MAC protocol is to apply the RTS/CTS/DATA/ACK four-way handshaking procedure and the DATA/ACK two-way handshaking procedure for long packets and for short packets, respectively. Based on this MAC protocol, if overhearing an RTS frame under the condition that no CTS frame is received, a node can establish another link to send a small-sized packet based on the DATA/ACK two-way handshaking procedure. With the NAV value in the overheard RTS frame of other nodes, the sender of the second link can schedule the transmission of the DATA frame to be synchronized with that of the first link. Similarly, the transmission time of the ACK frames in both the first link and the second link can be synchronized. The P-MAC protocol can achieve the objective of concurrent transmissions by simply not using RTS/CTS in the sending small-sized packets. Because approximately 50% of packets have a size smaller than 100 bytes in the Internet, the P-MAC protocol is quite suitable for delivering the traffic in the Internet.

To summarize, the P-MAC protocol can solve the virtual-carrier hidden node problem of Fig. 1. This MAC protocol can also overcome the physical-carrier exposed node problem of Figure 1 because the RTS/CTS handshaking mechanism is employed. Nevertheless, the virtual-carrier exposed node problem still cannot be alleviated by adopting the P-MAC protocol.

3 The proposed MAC protocol

In this section, we describe the proposed protocol in detail. We present a location-assisted media access control (MAC) protocol that exploits location information to validate potential concurrent transmissions. Such a scheduled transmission should not interfere with the current transmission, and it should not be corrupted by the current transmission. The scheduling of concurrent transmissions is transparent to the current sender-receiver pair, making it backward comparable with the original IEEE 802.11 MAC and S-MAC.

3.1 ACQUIRING LOCATING INFORMATION OF NODES

Nodes can get its location information, either through the Global Positioning System or many available localization algorithms. All nodes in the networks broadcast their location information, so each node can get location information for all neighbour nodes. Neighbour nodes add location information status field in the node list in order to store location information of node coordinates. Position in the node when that information is broadcast within the hop neighbour position, when the node location information changes, the node can update the neighbour list of neighbour location information value, so that each node will be saved within the two hop neighbour nodes.

3.2 CONFIRMING EXPOSED TERMINAL NODES

When a node first overhears an RTS and then a DATA frame from the same transmitter, it can be identified as an exposed terminal with regard to the overheard transmission. In practice, however, information such as frame type will only be available at the MAC layer after the entire frame is received and verified by checking the frame check sequence trailer.

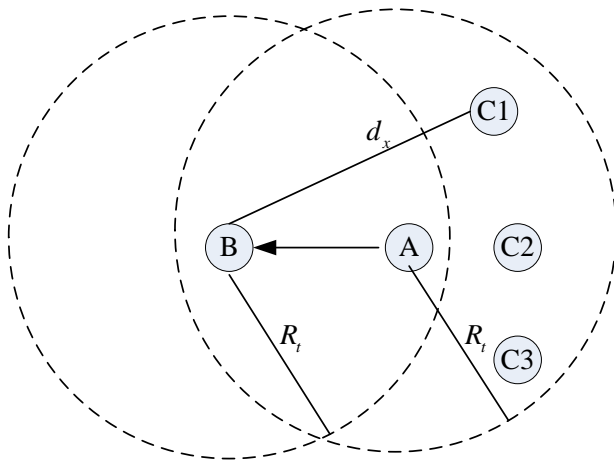


FIGURE 2 Identifying an Exposed Terminal

When a node recognizes itself as an exposed terminal in this manner, it will be too late to schedule any concurrent transmission. To schedule concurrent transmissions, exposed terminals should be identified before the current DATA frame transmission begins. At the same time according location information of nodes transmission distance can be calculated from the master node and receiving node. In Figure 2, if $d_x > R_t$, we can confirm C1, C2 and C3 are exposed terminals.

3.3 CHECKING CONCURRENT TRANSMISSION

All nodes are assumed to have the same type of radio and identical transmit power. We consider the two-ray ground

propagation model. Under this model, the relation between the transmit power P_t and the received power P_r is given by:

$$P_r = \frac{P_t G_t G_r (h_t h_r)^2}{d^4}, \tag{1}$$

where G_t and G_r are the gains of the transmit and receive antennas, respectively; h_t and h_r are the heights of transmit and receive antennas, respectively; and d is the distance between the transmitter and the receiver.

For a target receiver, let P_r denote the received transmit power and P_i denote the received interference power. For successful reception, the signal-to-interference ratio (SIR) at the receiver should be greater than a threshold SIR_T :

$$SIR = \frac{P_r}{P_i} \geq SIR_T > 1. \tag{2}$$

From the two-ray ground model and (1), the interference range R_i is defined as:

$$R_i = d_t^4 \sqrt{SIR_T}. \tag{3}$$

We focus on the successful transmission range when a scheduled transmission is allowed. Based on (2), boundary points of the area within which the scheduled transmission will not be interfered by the current transmission can be calculated from $n\sqrt{x^2 + y^2} \leq \sqrt{(x - d_0)^2 + y^2}$. Rearranging the aforementioned equation, we obtain:

$$\left(\frac{nd}{n^2 - 1}\right)^2 \leq \left(x + \frac{d_0}{n^2 - 1}\right)^2 + y^2. \tag{4}$$

That is, this region is a disk centered at $(-d_0/n^2 - 1, 0)$ with a radius $nd/(n^2 - 1)$. On the other hand, the scheduled receiver should be located within the scheduled transmitter's transmission range to correctly receive the frame, i.e.:

$$R_t^2 \geq x_2^2 + y_2^2. \tag{5}$$

Nodes that are out of the dashed circle cannot successfully decode the scheduled transmission, even if the concurrent transmission is absent.

3.4 EXECUTING CONCURRENT TRANSMISSION

By exploiting location information, we can correctly identify scheduled transmissions that will not be interfered by the concurrent transmission in Figure 3. Furthermore, for both DATA frames to be successfully delivered, the corresponding ACK transmissions should not interfere with each other either. In the case of ACK transmission, the roles of transmitter and receiver are switched, as shown in Figure 3. Therefore, to prevent collisions between the two ACK frames, the current receiver should be out of the interference range of the scheduled transmitter, and the scheduled receiver should be out of the interference range of the concurrent transmitter.

In Figure 3 d_{AD} is the distance between the scheduled transmitter and the current receiver, d_{BC} is the distance between the current transmitter and the scheduled receiver. These can easily be computed from the coordinates of these nodes. R_b and R_d are the interference ranges of the current receiver and the scheduled receiver, respectively. If $d_{AD} > R_b$ and $d_{BC} > R_d$ then exposed terminal C and D can successfully execute scheduled transmission which should not interfere with the current transmission, and it should not be corrupted by the current transmission.

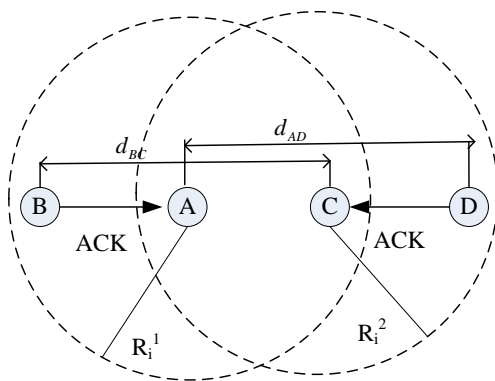


FIGURE 3 Transmitting ACK Frame when Concurrent Transmission

4 Simulation results

In this section, the performance of the proposed MAC protocol is evaluated via simulations and compared with S-AMC. Together with overhearing avoidance and message passing, S-MAC obtains significant energy savings compared with IEEE 802.11-like protocols without sleeping. In the simulation we take into account throughput and delay of the network. We suppose that each node has data frame to be constantly sent, i.e. the system is saturated.

In our work, we simulate network throughput and transmission delay by changing the network load. Firstly, we randomly distributed five scenarios with different number nodes without loss of generality in the network.

We consider a 1000m*1000m area, and put sink node in the centre of network region. Simulation parameter settings are as follows:

TABLE 1 Simulation Parameter Settings

Parameter statement	value	Parameter statement	value
Data frame length	1024bytes	Channel rate	1MB/s
MAC heads length	224bits	SIFS interval	10us
ACK frame length	304bits	DIFS frame interval	50us
RTS frame length	352bits	Minimum contention window	31
CTS frame length	304bits	Maximum contention window	1023
PHY frame length	192bits	Slot length	20us

Figure 4 shows the analysis of end-to-end throughput with five different scenarios. We can see that with increasing the number of nodes in the network, the proposed MAC achieve the more chance for parallel transmission of exposed terminal and the end to end throughput increase too. However, due to S-MAC with carrier sensing mechanism prevents nodes from the parallel transmission, so only one of them can access the channel at any time. As shown in Fig.4, there is not much change in the average throughput of S-MAC. Experimental results show that compared with S-MAC, the throughput of the proposed protocol has significantly improved. Therefore, the proposed protocol can better solve the traditional problem of low throughput of MAC protocol.

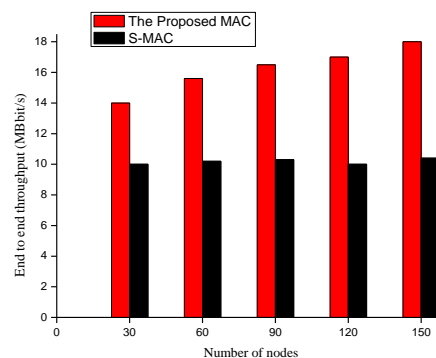


FIGURE 4 The analysis of end to end throughput with five different scenarios

Figure 5 shows the end-to-end delay with five different scenarios. We can see there is only one effective link for transmission of S-MAC, with increasing the number of nodes in the network, channel competition becomes fierce. S-MAC can get the channel after times of group competition, meanwhile, the networks increase the number of accumulated packets the conflicts increase as well, which led high retransmission rate, and packet transmission delay is in high state. Adopting the proposed MAC, exposed terminals can transmit concurrently so the end to end delay can be decreased largely. We find in

Figure 5 that the ratios of the end-to-end delay with the proposed protocol to that with the S-MAC are 64.28%, 60.6%, 70%, 80% and 78.57%. The proposed MAC works well in multi-hop wireless networks.

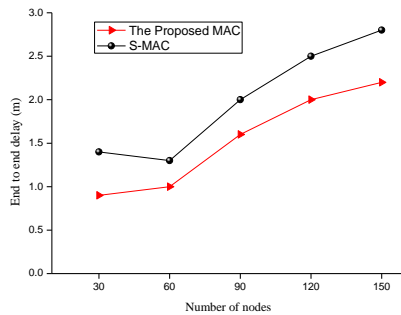


FIGURE 5 The analysis of end to end delay with five different scenarios

Secondly, the simulation study is performed with the chain network topology, where the distance between any two adjacent nodes is set to 200m. In the simulation, the forward flow (from Node 1 to Node N) is a constant-bit-rate (CBR) session with a stream of 1000-byte frames, whereas the backward flow (from Node N to Node 1) is a CBR session with a smaller packet size.

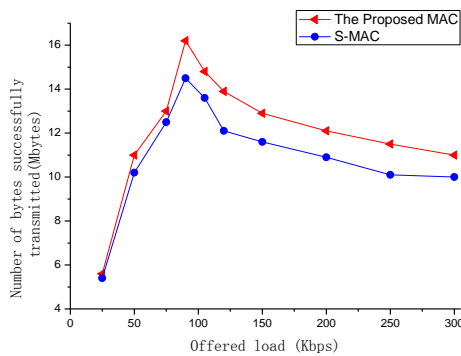


FIGURE 6 Throughput versus with offered load

We first plot the throughput versus offered load for an ten node chain network in Figure 6. For both schemes, we find that the throughput first increases with offered load in the under load region, due to the more data available for transmission. In the overload region, however, the throughput decreases with offered load due to congestion. We find that both MACs achieve the largest throughput when offered load is 90 kb/s.

Figure 7 shows the simulation results for chain networks with various packet sizes for the backward flow. Here, the data rates of the flows for the 8-, 10-, 12-, 14-, 16-node chain networks are set to 100kb/s, 90kb/s, 80kb/s, 70kb/s and 55kb/s, respectively. For the 8-,10-,12-,14-,16-node chain networks, the normalized improvements achieved by the proposed scheme are found to be 56.78%, 32.13%, 37.26% and 39.67%, respectively. When the packet size of the backward flow is reduced, the improvement ratio tends to be smaller. For example, when the packet size of the backward flow is 550 bytes,

the throughput improvement ratio ranges from 26.35% to 63.47% for chain networks with increasing number of nodes. When the packet size of the backward flow is 300 bytes, the throughput improvement ratio ranges from 15.28% to 58.39% for the chain networks.

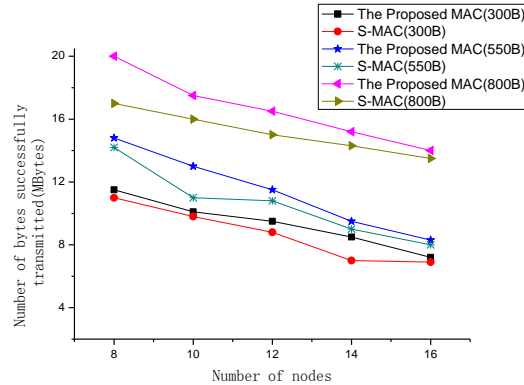


FIGURE 7 Throughput versus various packet sizes for backward flow

In Figure 8, we observe significantly reduced average end-to-end delays when the proposed MAC is used. Significant delay improvement is also achieved for other chain networks we examined, when the packet size of the backward flow is equal to 800 bytes for the 8-, 10-, 12-, 14- and 16-node networks, respectively.

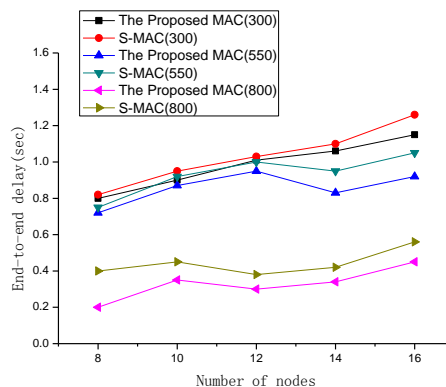


FIGURE 8 End-to-end delay versus various packet sizes for backward Flow

5 Conclusions



In wireless sensor network environment, it has been shown that traditional CSMA media access control suffers from a low-throughput problem, which is largely due to the inefficiency in carrier sensing and spatial reuse. In this paper, we present a location-assisted MAC protocol that schedules “feasible” concurrent transmissions in wireless sensor network. A simple procedure based on location information is adopted in the proposed MAC to validate the feasibility of a concurrent transmission. Our simulation results show that the proposed scheme can effectively increase the throughput and reduce the average end-to-end delay of wireless sensor networks.

Acknowledgments

This work is supported by Natural Science Foundation of Heilongjiang Province (F201104), Heilongjiang Province Education Department Foundation (12521416) and the Innovation Foundation of Harbin Bureau of Science and Technology (2011RFDXG005).

References

- [1] IEEE 802.11 Standard. *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications* New York: IEEE, 1999
- [2] Shukla D, Chandran-Wadia L, Iyer S 2003 Mitigating the exposed node problem in IEEE 802.11 ad hoc networks *Proc. IEEE Int. Conf. Comput. Commun. Netw.* Dallas TX, Oct 2003, 157–62
- [3] Reigadas S J, Martinez-Fernandez A, Seoane-Pascual J 2010 *IEEE Transactions on Mobile Computing* 9(6) 881-96
- [4] Guo W, Xiaofeng Z, Jing W 2011 A New Constrained-send Mechanism to Enhance the Performance of IEEE 802.11DCF *The 6th International ICST Conference 2011 China* 448-52
- [5] Yayu G, Xinghua S, Lin D 2013 *IEEE Transactions on Wireless Communications* 12(1) 398-411
- [6] Akcan H, Kriakov V, Brnnimann H, Delis A 2006 GPS-free node localization in mobile wireless sensor networks *Proc. 5th Int. ACM Workshop Data Chicago* 35-42
- [7] Shi Q, He C, Chen H, Jiang L 2010 *IEEE Transactions on Signal Processing* 58(6) 3328-40
- [8] Bencini L, Fantacci R, Maccari L 2010 Analytical Model for Performance Analysis of IEEE 802.11 DCF Mechanism in Multi-Radio Wireless Networks *Proc IEEE ICC 2010 23-27 Maggio 2010 No 7 Cape Town: South Africa* 1-5
- [9] Son I K, Mao S, Hur S M 2009 *Sensors* 9(6) 4824–44
- [10] Kim T-S, Lim H, Hou J C 2008 *IEEE Transactions on Mobile Computing* 7(10) 1200–12
- [11] Fu L, Liew S C, Huang J 2010 Effective Carrier Sensing in CSMA Networks under Cumulative Interference. *Proceedings of INFOCOM San Diego USA* 1-9
- [12] Jiang L, Walrand J 2010 *IEEE/ACM Trans on Networking* 18(3) 960-972
- [13] Acharya A, Misra A, Bansal S 2003 MACA-P: A MAC for concurrent transmissions in multi-hop wireless networks *Proc. IEEE Int. Conf. Pervasive Comput. Commun.* Dallas TX Mar. Hong Kong 505–8
- [14] Felemban E, Ekici E 2010 *IEEE Transactions on Wireless Communications* 10(10) 3256-66
- [15] Jingsung L, Hojin Lee, Yung Y, Song C, Nardelli B 2013 Making 802.11 DCF near-optimal: Design, implementation, and evaluation *The 10th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks USA* 86-94
- [16] Aydogdu C, Karasan E 2011 *IEEE Transactions on Mobile Computing* 10(10) 1361-73
- [17] Zhou Y H, Nettles S M 2005 Balancing the hidden and exposed node problems with power control in CSMA/CA-based wireless networks *IEEE Wireless Communications and Networking Conference New Orleans LA USA* 683-8
- [18] Shukla D, Chandran-Wadia L, Iyer S 2003 Mitigating the Exposed Node Problem in IEEE 802.11 Ad Hoc Networks *The 12th International Conference on Computer Communications and Networks* 157–62

Authors	
	<p>Guoyan Yang</p> <p>Current position, grades: a full Associate Professor of College of Information and Technology, Heilongjiang University University studies: M.Sc.(2007)and Ph.D.(2012) in computer science and technology (2007) from Harbin Engineering University Scientific interest: research interests include topology control, performance evaluation and routing algorithm in wireless sensor networks</p>
	<p>Xin Guan</p> <p>Current position, grades: a full Associate Professor of College of Information and Technology, Heilongjiang University. University studies: the Bachelor degree in School of Computer Science and Technology from Heilongjiang University, China in 2001 and the Master degree from Harbin Institute of Technology, China in 2007, the Ph.D. at graduate school of science and technology, Keio University, Japan (2012) Scientific interest: research interests include topology control, performance evaluation and routing algorithm in wireless sensor networks.</p>

Flow-insensitive type qualifier inference on programming languages allowing type casts

Huisong Li^{1, 2*}

¹State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, 100190, China

²University of Chinese Academy of Sciences, Beijing 100049, China

Received 1 January 2014, www.tsi.lv

Abstract

Type qualifiers are lightweight specifications of atomic properties that refine the standard types. Flow-insensitive type qualifier inference has been used in the CQual framework to improve the quality of C programs. However, type casts will affect the effectiveness of type qualifier inference, as they can lead to either accepting some flawed programs due to discarding some useful qualifier information, or rejecting some safe programs when the analysis is conservative. In this paper, we first present a language, which allows type casts and formalize its flow-insensitive qualifier inference system. We then show some examples to illustrate how qualifiers are lost because of type casts in CQual and give an idea on solving this problem.

Keywords: Type cast, Type inference, Flow-insensitive type

1 Introduction

Type system is one of the popular formal methods to express general correctness properties of programs [6]. Type systems can be considered as tools for reasoning about programs. For example, they can statically check whether the arguments of primitive arithmetic operations are always integers. However, type system cannot capture arbitrary program behaviour. They can only guarantee that a well-typed program is free from certain kinds of bad program behaviours. For instance, type systems cannot check that the second argument of the division operation is non-zero. Nevertheless, several methods have been investigated to extend the expressivity of type systems: generalized algebraic data types [19], dependent types [20, 21], refinement types [2, 12], type qualifiers [8] and so on, among which, only type qualifiers are lightweight.

Type qualifiers are lightweight specifications for specifying atomic properties that refine standard types. Generally, qualifiers fall into two classes; value qualifiers, such as **tainted** and **untainted**, and reference qualifiers, such as **const** and **nonconst**. Value qualifiers pertain only to the value of an expression and reference qualifiers pertain to the address of an expression [5].

Specifically, **tainted** and **untainted** are qualifiers specified for checking the format-string vulnerabilities of C programs [16]. Usually, data read from the environment which could be controlled by untrusted users, like the network users, should be annotated with **tainted**. Printing functions, like *printf*, require the first argument to be **untainted** as the format specifier. It is

safe to interpret **untainted** data as **tainted** data, but not vice versa, which can be represented as $\text{untainted} \leq \text{tainted}$. Type qualifiers **const** and **nonconst** are used in **const** inference [10]. Usually, a **nonconst** cell can be reassigned, but a **const** cell cannot be reassigned, so it is safe to interpret a nonconst cell as a const cell, but not vice versa, which can be represented as $\text{nonconst} \leq \text{const}$. In [10], Foster proposed a framework using type qualifiers to improve the quality of programs. In his framework, programmers are allowed to specify qualifiers and add qualifier annotations to the programs. Type qualifier inference, including flow-insensitive inference and flow-sensitive inference, will determine the remaining qualifiers and check the consistency of the qualifier annotations. In flow-insensitive inference system, a value's qualified type is the same everywhere, but always changes with the assignment in a flow-sensitive inference system. In both systems, however, the underlying types keep unchanged. Based on this framework, CQual was developed for C programs and has been used in a number of different practical applications [1, 11, 14, 16, 18].

Example 1 Example Program fragment in C:

```
const int* x;
int* y;
int a, b;
a = (int)x;           /*(1)*/
b = a;                /*(2)*/
y = (int*)b;          /*(3)*/
*y = 5;               /*(4)*/
```

However, type casts will affect the effectiveness of type qualifier inference, thus leading to missing some

* Corresponding author - E-mail: lihs@ios.ac.cn

flaws of programs. For the flow-insensitive inference, consider the example program fragment in Example 1, which can be accepted by CQual. In the program, x is a pointer to **const**. From the lines (1), (2), (3), x and y have the same value, which means y should be inferred as a pointer to **const** and the assignment should not be allowed at line (4). However, the type information that x is a pointer to **const** cannot be transferred to y by a and b because their type `int` cannot “contain” enough qualifier information as `int*`, which leads to losing some useful information during the qualifier inference process. However, note that at line (1), x is assigned to a after the type cast, which means the value of a is actually of type `int*`. The value of b is also of type `int*` as b is assigned to the value of a at line (2). Then, in the qualifier inference process, we can use `int*` instead of `int` for a and b . By doing so, we get that y is a pointer to **const**, that is, `int*` is more suitable for the qualifier inference process than `int` for a and b . Thus, for the qualifier inference process, we believe that the most suitable type of each expression may not be the type offered in the program.

The rest of this paper is organized as follows: Section 2 shows our source language and some preliminaries; Section 3 presents a flow-insensitive qualifier inference system for the source language; Section 4 shows more examples about losing qualifier information because of type casts in CQual and a simple idea on how to solve this problem. Section 5 presents the related work.

2 Preliminaries

In the paper, we present the basic theory using a simple call-by-value source language. Prior to introducing expressions, we first present the definitions of types and qualified types respectively. A type t is a term generated by the following grammar:

$$t ::= \text{int} \mid \text{ref}(t) \mid t \rightarrow t \quad ,$$

where `ref` is a pointer type constructor and \rightarrow is a function type constructor. Given a set of qualifiers Q , a qualified type τ is a term generated by:

$$\begin{aligned} \tau &::= q_c \sigma \quad q_c \in Q \\ \sigma &::= \text{int} \mid \text{ref}(\tau) \mid \tau \rightarrow \tau \quad . \end{aligned}$$

Definition 1 gives the definition of the expressions of the simple language, where $(t)e$ is a type cast expression that casts the type of expression e to type t and $\text{annot}(e, q_c)$ is a qualifier annotation expression, which specifies q_c as the outermost qualifier of the qualified type of e . Generally, a set of qualifiers Q and their order \leq form a partial order set. For simplicity, the partial order: $(Q, \leq) = \{\text{untainted} \leq \text{tainted}\}$ is used for the examples of the rest paper.

Example 2 Main Example Program:

```
let x = ref annot(0, tainted) in
let y = (int)x in
let z = (ref(int))y in
      annot(*z, untainted)
```

- *Definition 1* Source language with type casts and qualifier annotations:

$e ::=$	v	value
	$ e_1 e_2$	application
	$ \text{let } x = e_1 \text{ in } e_2$	bind
	$ \text{ref } e$	reference
	$ *e$	deference
	$ (t)e$	type cast
	$ e_1 := e_2$	assignment
	$ \text{annot}(e, q_c)$	annotation
$v ::=$	n	integer
	$ x$	variable
	$ \lambda x : t. e$	function

Example 2 shows a program, which is written in our simple language and shows type casts between pointer type `ref(int)` and `int`. This program is not safe as `*x` and `*z` refer to the same object, which means we interpret a **tainted** object as an **untainted** object. Therefore, we expect to find this flaw by the qualifier inference system.

The partial order on type qualifiers can induce the qualified subtype relation \leq (In this paper, we use \leq for both the order between qualifiers and qualified types.) among qualified types in Definition 2 [10]. $t_1 \triangleright t_2$ means it is entirely safe to interpret an object of type τ_1 as an object of type τ_2 . In order to ensure that all aliases of the same `ref` cell contain the same qualifiers, the second rule, (Ref_{\leq}) , requires that $\tau_1 = \tau_2$ ($\tau_1 \leq \tau_2$ and $\tau_2 \leq \tau_1$) instead of $\tau_1 \leq \tau_2$. Otherwise, there will be problems such as assigning a pointer of type `ref(untainted int)` to pointer of type `ref(tainted int)` where we can write **tainted** data to **untainted** position through the `ref(tainted int)` pointer. In the rule, $\text{strip}_q(\tau_1) \triangleright \text{strip}_q(\tau_2)$, functions are contravariant in their domain and covariant in their range. In Definition 3, embed_q is a function for mapping standard types to qualified types with fresh qualifier variables. Qualifier variables (written as q) stand for unknown qualifiers. strip_q is a function mapping qualified types to their standard types.

The type casts allowed in our source language are showed in Definition 4. If the cast from t_1 to t_2 is allowed, we say t_1 is the castable type of t_2 , denoted as $t_1 \triangleright t_2$, meaning that a value of type t_1 can be interpreted

as a value of type t_2 . ($\text{IR}_{\triangleright}$) and ($\text{RI}_{\triangleright}$) show all type casts between the pointer type and type int are allowed as an integer can be interpreted as an address and vice versa. ($\text{RR}_{\triangleright}$) shows all type casts between pointer types are allowed as an address of one type can be interpreted as an address of another type. For simplicity, we forbid all the other type casts, like type casts between function types.

Based on the castable type relation, we define the following qualified castable type relation. We first give the rules of the relation \triangleright^q in Definition 5, in which $\tau_1 \triangleright^q \tau_2$ means their “corresponding” qualifiers satisfy the partial order \leq of (Q, \leq) . Rules, ($\text{RR}_{\triangleright^q}$) and ($\text{FF}_{\triangleright^q}$), are similar to the rules of the qualified subtype relation. All the other rules only have the outermost qualifiers as the “corresponding” qualifiers. Note that the rules in Definition 5 are general formalizations for value qualifiers, like **tainted** and **untainted**. For the other qualifier kinds, the rules can be adjusted according to the properties of qualifiers and our analysis requirements. If $\text{strip}_q(\tau_1) \triangleright \text{strip}_q(\tau_2)$ (the standard type of τ_1 is the castable type of the standard type of τ_2) and $\tau_1 \triangleright^q \tau_2$, then τ_1 is the qualified castable type of τ_2 , meaning that interpreting a value of type τ_1 as a value of type τ_2 is partially safe. The reason is that τ_1 is the qualified castable type of τ_2 and τ_2 is the qualified castable type of τ_3 does not guarantee that τ_1 is the qualified castable type of τ_3 .

- *Definition 2* Qualified subtype relation:

$$\frac{q_1 \leq q_2}{q_1 \text{ int } \leq q_2 \text{ int}} \quad (\text{Int}_{\leq})$$

$$\frac{q_1 \leq q_2 \quad \tau_1 = \tau_2}{q_1 \text{ ref}(\tau_1) \leq q_2 \text{ ref}(\tau_2)} \quad (\text{Ref}_{\leq})$$

$$\frac{q_1 \leq q_2 \quad \tau_3 \leq \tau_1 \quad \tau_2 \leq \tau_4}{q_1 (\tau_1 \rightarrow \tau_2) \leq q_2 (\tau_3 \rightarrow \tau_4)} \quad (\text{Fun}_{\leq})$$

- *Definition 3* embed_q and strip_q :

$$\begin{aligned} \text{embed}_q(\text{int}) &= q \text{ int} \quad q \text{ fresh} \\ \text{embed}_q(\text{ref}(t)) &= q \text{ ref}(\text{embed}_q(t)) \quad q \text{ fresh} \\ \text{embed}_q(t_1 \rightarrow t_2) &= \\ & q (\text{embed}_q(t_1) \rightarrow \text{embed}_q(t_2)) \quad q \text{ fresh} \\ \text{strip}_q(q \text{ int}) &= \text{int} \\ \text{strip}_q(q \text{ ref}(\tau)) &= \text{ref}(\text{strip}_q(\tau)) \\ \text{strip}_q(q (\tau_1 \rightarrow \tau_2)) &= \text{strip}_q(\tau_1) \rightarrow \text{strip}_q(\tau_2) \end{aligned}$$

- *Definition 4* Castable-type relation:

$$\begin{aligned} \text{int} \triangleright \text{int} & \quad (\text{II}_{\triangleright}) \\ \text{int} \triangleright \text{ref}(t) & \quad (\text{IR}_{\triangleright}) \\ \text{ref}(t) \triangleright \text{int} & \quad (\text{RI}_{\triangleright}) \\ \text{ref}(t_1) \triangleright \text{ref}(t_2) & \quad (\text{RR}_{\triangleright}) \end{aligned}$$

- *Definition 5* Qualified castable-type relation:

$$\frac{q_1 \leq q_2}{q_1 \text{ int } \triangleright^q q_2 \text{ int}} \quad (\text{II}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 \text{ int } \triangleright^q q_2 \text{ ref}(\tau)} \quad (\text{IR}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 \text{ ref}(\tau) \triangleright^q q_2 \text{ int}} \quad (\text{RI}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 \text{ int } \triangleright^q q_2 (\tau_1 \rightarrow \tau_2)} \quad (\text{IF}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 (\tau_1 \rightarrow \tau_2) \triangleright^q q_2 \text{ int}} \quad (\text{FI}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 \text{ ref}(\tau_1) \triangleright^q q_2 (\tau_2 \rightarrow \tau_3)} \quad (\text{RF}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2}{q_1 (\tau_1 \rightarrow \tau_2) \triangleright^q q_2 \text{ ref}(\tau_3)} \quad (\text{FR}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2 \quad \tau_1 \triangleright^q \tau_2 \quad \tau_2 \triangleright^q \tau_1}{q_1 \text{ ref}(\tau_1) \triangleright^q q_2 \text{ ref}(\tau_2)} \quad (\text{RR}_{\triangleright^q})$$

$$\frac{q_1 \leq q_2 \quad \tau_3 \triangleright^q \tau_1 \quad \tau_2 \triangleright^q \tau_4}{q_1 (\tau_1 \rightarrow \tau_2) \triangleright^q q_2 (\tau_3 \rightarrow \tau_4)} \quad (\text{FF}_{\triangleright^q})$$

3 The Qualifier Inference System

For the source language, the standard type checking system is presented in Definition 6. Judgements of the form $\Gamma \vdash e : t$ mean that in the type environment Γ which maps variables to their types, expression e has type t . (Cast) computes the type t' of e , checks that t' is the castable type of t , and then t is the type of the **cast** expression. (Annot) ignores the annotation of the qualifier as standard types do not contain qualifiers. The type t of e is the type of the **Annot** expression. The discussion of the other rules can be found in [10]. Note that the standard static type checking system cannot guarantee that programs run without runtime errors because of type casts, which is a well-known limitation of type systems.

The qualifier inference system in Definition 7 is used to infer qualifiers automatically under the assumption that the program passes the standard type checking system. For this system, judgements of the form $\Gamma_q \vdash e : \tau$ mean that in qualified type environment Γ_q , which maps variables to their qualified types, expression e has qualified type τ . Whenever we assign a type to a term constructor, we introduce a fresh qualifier variable to stand for the unknown qualifier on the term that we need to solve for (see (Int_q), (Lam_q) and (Ref_q)). We use embed_q in (Lam_q) and (Cast_q) to map the given standard type to a qualified type with fresh qualifier variables. Rules (App_q) and (Assign_q) generate the qualified subtype constraints of the form $\tau_1 \leq \tau_2$. (Cast_q) generates the qualified castable type constraints of the form $\tau_1 \triangleright^q \tau_2$. (Annot_q) computes the qualified type $q \sigma$ of e and generates the constraints

$q = q_c$ ($q \leq q_c$ and $q_c \leq q$) in order to infer that the value of q is q_c .

After we perform the qualifier inference system on the program, we will get a set C_τ containing constraints of the forms $\tau_1 \leq \tau_2$, $\tau_1 \triangleright^q \tau_2$ and a set C_q containing constraints of the form $q = q_c$. Then, we reduce all constraints of C_τ into C_q using $R(\tau, \tau')$ in Definition 8 which are simply the rules of Definition 5 rewritten left-to-right.

Formally,

$$C_q = C_q \cup \bigcup_{\forall \tau_1 \leq \tau_2 \in C_\tau} R(\tau_1, \tau_2) \cup \bigcup_{\forall \tau_1 \triangleright^q \tau_2 \in C_\tau} R(\tau_1, \tau_2).$$

The algorithm for solving C_q to get the value of qualifier variables can be found in [10]. Note that as we assume the program is correct with respect to the standard type checking system, then for any $\tau_1 \leq \tau_2 \in C_\tau$, $\text{strip}_q(\tau_1) = \text{strip}_q(\tau_2)$ (syntactic equality) and for any $\tau_1 \triangleright^q \tau_2 \in C_\tau$, $\text{strip}_q(\tau_1) \triangleright \text{strip}_q(\tau_2)$.

Example 3 The inference result of the Example 2:

$$\Gamma_q : \left\{ \begin{array}{l} (x, q_2 \text{ ref } (q_1 \text{ int})), (y, q_3 \text{ int}), \\ (z, q_5 \text{ ref } (q_4 \text{ int})) \end{array} \right\}$$

$$C_\tau : \left\{ \begin{array}{l} q_2 \text{ ref } (q_1 \text{ int}) \triangleright^q q_3 \text{ int}, \\ q_3 \text{ int} \triangleright^q q_5 \text{ ref } (q_4 \text{ int}) \end{array} \right\}$$

$$C_q : \{q_1 = \mathbf{tainted}, q_4 = \mathbf{untainted}\}$$

Then, reducing C_τ into C_q will get that $C_q = \{q_1 = \mathbf{tainted}, q_4 = \mathbf{untainted}, q_2 \leq q_3, q_3 \leq q_5\}$. Solving C_q will get that the value of q_1 and q_4 are **tainted** and **untainted** respectively. However, as we have discussed in Example 2, this program is not safe by interpreting a **tainted** object as an **untainted** object. But this system does not find the flaw because the qualifier information of q_1 cannot be transferred to q_4 . That is, some qualifier information is lost during the inference process.

- *Definition 6* The standard type checking system:

$$\frac{x \in \text{dom}(\Gamma)}{\Gamma \text{ h } x : \Gamma(x)} \quad (\text{Var})$$

$$\frac{}{\Gamma \text{ h } n : \text{int}} \quad (\text{Int})$$

$$\frac{\Gamma[x \mapsto t] \text{ h } e : t'}{\Gamma \text{ h } \lambda x : t.e : t \rightarrow t'} \quad (\text{Lam})$$

$$\frac{\Gamma \text{ h } e_1 : t \rightarrow t' \quad \Gamma \text{ h } e_2 : t}{\Gamma \text{ h } e_1 e_2 : t'} \quad (\text{App})$$

$$\frac{\Gamma \text{ h } e : t}{\Gamma \text{ h } \text{ref}(e) : \text{ref}(t)} \quad (\text{Ref})$$

$$\frac{\Gamma \text{ h } e : \text{ref}(t)}{\Gamma \text{ h } *e : t} \quad (\text{Deref})$$

$$\frac{\Gamma \text{ h } e_1 : t \quad \Gamma[x \mapsto t] \text{ h } e_2 : t'}{\Gamma \text{ h } \text{let } x = e_1 \text{ in } e_2 : t'} \quad (\text{Let})$$

$$\frac{\Gamma \text{ h } e_1 : \text{ref}(t) \quad \Gamma \text{ h } e_2 : t}{\Gamma \text{ h } e_1 := e_2 : t} \quad (\text{Assign})$$

$$\frac{\Gamma \text{ h } e : t' \quad t' \triangleright t}{\Gamma \text{ h } (t)e : t} \quad (\text{Cast})$$

$$\frac{\Gamma \text{ h } e : t}{\Gamma \text{ h } \text{annot}(e, q_c) : t} \quad (\text{Annot})$$

4 Examples

Based on the inference of **tainted** and **untainted**, we list some examples in Table 1 to specifically show how CQual keep track of qualifiers in the presence type casts. For simplicity, we only show the source qualified types in the first column, the destination qualified types in the second column and the generated constraints for transferring the qualifier information of the source to the destination in the last column. The second line shows casting from $q_1 \text{ ref}(q_2 \text{ int})$ to $q_3 \text{ int}$. The constraint $q_1 \leq q_3$ transfers the qualifier information of q_1 to q_3 . Nevertheless, the qualifier information of q_2 is discarded. The forth line shows casting between pointers. From the qualified castable type relation, $q_2 = q_4$ ($q_2 \leq q_4$ and $q_4 \leq q_2$) is necessary. The last two lines show casting related to function pointers. We believe the constraints are some conservative, for example in the last line $q_2 = q_7$ will transfer the qualifier information of q_2 to q_7 . To conclude, we can get that there are always some qualifier information which may be useful been discarded. On the other side, some constraints on qualifiers are some conservative. These will affect the effectiveness of qualifier inference, thus leading to miss some flawed programs or produce some false positives.

Consider the second line of Table 1 again, if we augment $q_3 \text{ int}$ to $q_3 \text{ ref}(q_4 \text{ int})$ in which q_4 is a fresh

qualifier variable, we can preserve the qualifier information of q_2 to q_4 instead of discarding it directly. Note that the augmentation is only for the qualifier inference. However, for the type cast in the last line of Table 1, in order to preserve the qualifier information of q_3 and q_4 , we have to augment the destination type to contain a function type. The only way is introducing union types. We expect the augmentation result is $q_5 \text{ ref}(q_6 (\text{ref}(q_7 \text{ int}) \vee (q_8 \text{ int} \rightarrow q_9 \text{ int})))$, in which q_8 and q_9 are fresh qualifier variables. Then we can preserve the qualifier information of q_3 and q_4 to q_8 and q_9 respectively. We expect to do this in the future work.

5 Related work and conclusion

Type qualifiers can be seen as a kind of refinement types [2, 12], which do not change the underlying type structure and extend the expressivity of type systems. The type refinement framework in [15] supports a flow-sensitive, sophisticated type system. The theory of type qualifiers proposed in [10] describes a framework for adding type qualifiers to a language and show a flow-insensitive type qualifier inference system. Flow-sensitive type qualifier inference systems were proposed in [9], in which only type qualifiers were modelled flow-sensitively. In [5], a framework for allowing users to explicitly write type rules for their new type qualifiers and to explicitly specify the run-time invariant that the type qualifiers meant to represent was proposed, but it was not flexible enough, only supporting certain kinds of qualifiers. Based on the theory of type qualifiers, CQual [7] was developed for C programs and has had many applications, like const qualifiers inference[10], finding format-string vulnerabilities [16], static analysis of authorization hook placement [11, 18], finding user/kernel bugs [10], and deadlocks in Linux kernels [1]. Later, JQual [13] was developed for adding user-defined type qualifiers to Java.

Program analysis based on type qualifiers is a kind of static analysis techniques for improving software quality. However, as C standard allows arbitrary type casts between pointer types and some other type casts [17], we will lose some useful qualifier information during the qualifier inference process. Our work aims to solve this problem in the flow-insensitive qualifier inference process. Therefore, it can be seen as an improvement of flow-insensitive qualifier inference.

In this paper, we formalized a flow-insensitive qualifier inference system for a source language allowing type casts, and showed the problem of losing qualifier information caused by type casts in CQual and proposed a simple idea to solve this problem.

- *Definition 7* Qualifier inference system:

$$\frac{x \in \text{dom}(\Gamma_q)}{\Gamma_q \text{ h}_q x : \Gamma_q(x)} \quad (\text{Var}_q)$$

$$\frac{q \text{ fresh}}{\Gamma_q \text{ h}_q n : q \text{ int}} \quad (\text{Int}_q)$$

$$\frac{\tau = \text{embed}_q(t) \quad \Gamma_q[x \mapsto \tau] \text{ h}_q e : \tau'}{\Gamma_q \text{ h}_q \lambda x : t.e : q (\tau \rightarrow \tau')} \quad (\text{Lam}_q)$$

$$\frac{\Gamma_q \text{ h}_q e_1 : q (\tau_1 \rightarrow \tau_2) \quad \Gamma_q \text{ h}_q e_2 : \tau_3 \quad \tau_3 \leq \tau_1}{\Gamma_q \text{ h}_q e_1 e_2 : \tau_2} \quad (\text{App}_q)$$

$$\frac{\Gamma_q \text{ h}_q e : \tau \quad q \text{ fresh}}{\Gamma_q \text{ h}_q \text{ ref}(e) : q \text{ ref}(\tau)} \quad (\text{Ref}_q)$$

$$\frac{\Gamma_q \text{ h}_q e : q \text{ ref}(\tau)}{\Gamma_q \text{ h}_q *e : \tau} \quad (\text{Deref}_q)$$

$$\frac{\Gamma_q \text{ h}_q e_1 : \tau \quad \Gamma_q[x \mapsto \tau] \text{ h}_q e_2 : \tau'}{\Gamma_q \text{ h}_q \text{ let } x = e_1 \text{ in } e_2 : \tau'} \quad (\text{Let}_q)$$

$$\frac{\Gamma_q \text{ h}_q e_1 : q \text{ ref}(\tau) \quad \Gamma_q \text{ h}_q e_2 : \tau' \quad \tau' \leq \tau}{\Gamma_q \text{ h}_q e_1 := e_2 : \tau'} \quad (\text{Assign}_q)$$

$$\frac{\Gamma_q \text{ h}_q e : \tau \quad \tau' = \text{embed}_q(t) \quad \tau \triangleright^q \tau'}{\Gamma_q \text{ h}_q (t)e : \tau'} \quad (\text{Cast}_q)$$

$$\frac{\Gamma_q \text{ h}_q e : q \sigma \quad q = q_c}{\Gamma_q \text{ h}_q \text{ annot}(e, q_c) : q \sigma} \quad (\text{Annot}_q)$$

- *Definition 8* Constraint resolution rules:

$$R(q_1 \text{ int}, q_2 \text{ int}) = \{q_1 \leq q_2\}$$

$$R(q_1 \text{ int}, q_2(\tau_1 \rightarrow \tau_2)) = \{q_1 \leq q_2\}$$

$$R(q_1 \text{ int}, q_2 \text{ ref}(\tau)) = \{q_1 \leq q_2\}$$

$$R(q_1(\tau_1 \rightarrow \tau_2), q_2 \text{ int}) = \{q_1 \leq q_2\}$$

$$R(q_1 \text{ ref}(\tau), q_2 \text{ int}) = \{q_1 \leq q_2\}$$

$$R(q_1 \text{ ref}(\tau_1), q_2(\tau_2 \rightarrow \tau_3)) = \{q_1 \leq q_2\}$$

$$R(q_1 \text{ ref}(\tau_1), q_2 \text{ ref}(\tau_2)) = \{q_1 \leq q_2\} \cup R(\tau_1, \tau_2) \cup R(\tau_2, \tau_1)$$

$$R(q_1(\tau_1 \rightarrow \tau_2), q_2 \text{ ref}(\tau_3)) = \{q_1 \leq q_2\}$$

$$R(q_1(\tau_1 \rightarrow \tau_2), q_2(\tau_3 \rightarrow \tau_4)) = \{q_1 \leq q_2\} \cup R(\tau_3, \tau_1) \cup R(\tau_2, \tau_4)$$

References

- [1] Aiken A, Foster J S, Kodumal J, Terauchi T 2003 *ACM SIGPLAN Notices* 38(5) 129–40
- [2] Bierman G M, Gordon A D, Hrițcu C, Langworthy D 2010 *ACM SIGPLAN Notices* 45(9) 105–16
- [3] Castagna G, Xu Zhiwu. 2011 *ACM SIGPLAN Notices* 46(9) 94–106
- [4] Chandra S, Reps T 1999 *ACM SIGSOFT Software Engineering Notes* 24(5) 66–75
- [5] Chin B, Markstrum S, Millstein T 2005 *ACM SIGPLAN Notices* 40(6) 85–95
- [6] Pierce B C 2002 *Types and programming languages* MIT Press
- [7] Foster J S 2001 *CQual User's Guide*. Berkeley: University of California
- [8] Foster J S, Fähndrich M, Aiken A 1999 *ACM SIGPLAN Notices* 34(5) 192–203
- [9] Foster J S, Terauchi T, Aiken A 2002 *ACM SIGPLAN Notices* 37(5) 1–12
- [10] Foster J S 2002 *Type qualifiers: lightweight specifications to improve software quality* Berkeley: University of California

- [11]Fraser T, Petroni Jr N L, Arbaugh W A 2006 Applying flow-sensitive CQUAL to verify MINIX authorization check placement *Proceedings of the workshop on Programming languages and analysis for security* ACM 3-6
- [12]Freeman T, Pfenning F 1991 *ACM SIGPLAN Notices* 26(6) 268-77
- [13]Greenfieldboyce D, Foster J S 2007 *ACM SIGPLAN Notices* 42(10) 321-36
- [14]Johnson R, Wagner D 2004 Finding User/Kernel Bugs With Type Inference *Proceedings of the 13th Usenix Security Symposium, San Diego, CA*
- [15]Mandelbaum Y, Walker D, Harper R 2003 *ACM SIGPLAN Notices* 38(9) 213-25
- [16]Shankar U, Talwar K, Foster J S, et al 2001 Detecting Format String Vulnerabilities with Type Qualifiers *USENIX Security Symposium* 201-20
- [17]Siff M, Chandra S, Ball T, et al. 1999 Coping with type casts in C. *Software Engineering—ESEC/FSE '99*. Springer Berlin Heidelberg: 180-98
- [18]Zhang Xiaolan, Edwards A, Jaeger T 2002 Using CQUAL for Static Analysis of Authorization Hook Placement. *USENIX Security Symposium* 33-48
- [19]Xi Hongwei, Chiyan Chen, Gang Chen 2003 *ACM SIGPLAN Notices* 38(1) 224-35
- [20]Augustsson L 1999 Cayenne - a language with dependent types *Advanced Functional Programming* Berlin-Heidelberg: Springer 240-267
- [21]Xi Hongwei, Pfenning F 1999 Dependent types in practical programming *POPL '99 Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of programming languages* 214-227

Author

**Huisong Li****University studies:** master student in University of Chinese Academy of Sciences**Scientific interest:** research interests are programming language and static analysis

Fragments based tracking with adaptive multi-cue integration

Lichuan Gu*, Jianxiao Liu, Chengji Wang

School of Computer and Information, Anhui Agriculture University, Hefei Anhui, No.130 West Changjiang Road China, 230036

Received 1 January 2014, www.tsi.lv

Abstract

In this paper, we address the issue of part-based tracking by proposing a new fragments-based tracker with multi-cue integration. First the target template is divided into multiple fragments to get the contribution of each fragment to the possible positions of the target in the current frame, similarity measure is used in edge histogram and HSV histogram in every fragment, and the weights of cues integration are computed adaptively. Second we present a fragment template update mechanism with the discrimination between occlusions and appearance changes. The template is unchanged when the target is occluded and some fragments of template are updated in the case of appearance changes. In the experiments we use the indoor and outdoor test videos which contain the illumination changes, occlusions, and the appearance changes of targets. The experimental results show that our approach has strong robustness and high tracking accuracy.

Keywords: Target tracking, Multi-cue integration, Fragments tracking, Template update

1 Introduction

The object tracking in an accurate way is essential for applications like activity analysis, man-machine interaction and visual surveillance. However, in the actual process of tracking, target appearance changes (as shown in Figure 1 (a)) and occlusions (as shown in Figure 1 (b)) will affect the stability of the tracking algorithm, which leads to inaccurate target tracking and even target lost. As a result, how to effectively deal with target appearance changes and occlusions has always been one of the difficulties in target tracking problems.

A common solution to the above problems is tracking target by the integration of multiple visual cues, namely each cue provides a likelihood value for possible positions of the target in the next frame, and these cues are integrated according to the likelihood values to locate the final output (Figure 2). In recent ten years, a few of tracking algorithms using multi-cue integration have been proposed, say, Birchfield proposed using intensity gradient and color histograms to track people head [1]. The disadvantage of this method is that during tracking the cues do not always provide reliable information about the target object, since the equal weights are assigned to each cue and the cues' weights are fixed in the whole video sequence. In order to solve this problem, Triesch and von der Malsburg proposed a new integration framework [2], which can take advantage of the uncertainty of each cue to adaptively adjust each cue contribution to the tracking results. Following this framework appeared many adaptive cues integration algorithms, as shown in [3-8]. These algorithms integrated a variety of cues, complemented each other

and improved the accuracy of the tracking process to some extent.



FIGURE 1 Difficulties in tracking conditions: (a) target's variable appearance; (b) occlusions

Another solution to the above problems is to track moving targets based on fragments. In the tracking of

* Corresponding author - E-mail: glc@ahau.edu.cn

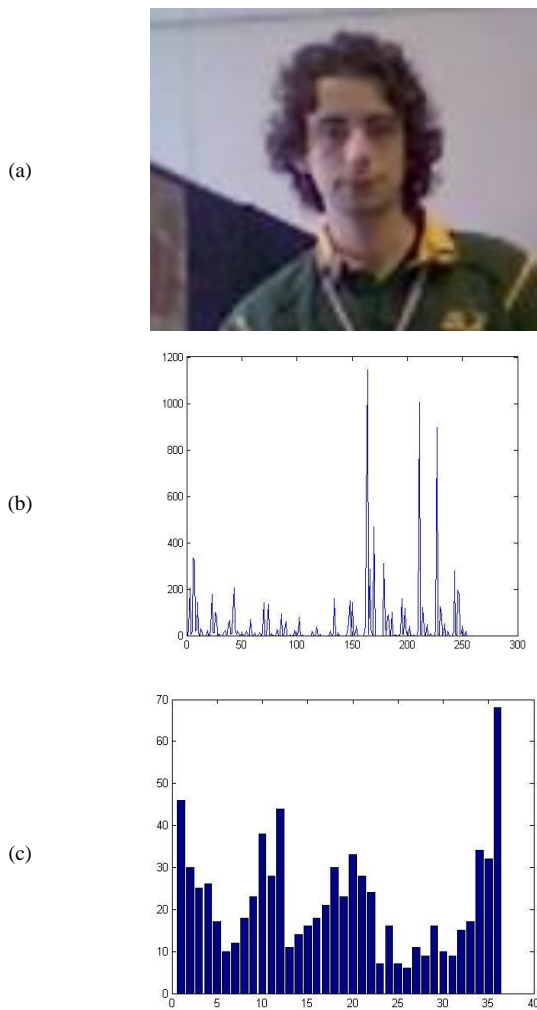


FIGURE 2 Cues based on different features: (a) Original image; (b) HSV histogram; (c) Edge histogram

human, for example, human is divided into the head, torso and limbs [9-11]. This approach generally required target model is known or a priori premise and is not suitable for generic target tracking. In order to solve this problem, some scholars put forward the general fragment tracking algorithms [12, 13]. This kind of method divided the target into different parts, but the division was arbitrary without considering any reference target model. In the process of tracking, this method weighted each fragment, and the contributions of different fragments are combined through statistical method to get the final output location of target. If there were target appearance changes or occlusions, the weight of corresponding fragment would be small and the impact on the overall goal would be small, thus the resolution of target would be improved.

The target tracking method based on template matching has attracted increasing attention [14-17]. Traditional image template matching based tracking method is still widely used because of its simple computations. This kind of method extracted some cues as a template which will be kept unchanged, and then

looked for a region whose cues were most similar to the template in current frame. However, the target may be occluded during motion, and may also change its appearance due to own motion. In these cases the template needs to be updated online to track the target accurately. The existing main template update algorithms update template once the changes of grey are detected in target regions. These algorithms are different from one another only in that the update rate or updated components are not the same, without discrimination of appearance changes and occlusions. The comparison of aforementioned algorithms is given in Table 1.

This paper's contributions lie in: 1) our algorithm employs an adaptive cue integration scheme in each fragment of the target. Similarity measure in edge histogram and HSV histogram are used, the vote of each fragment contributes to the joint tracking result according to adaptive integration, and the ones having small values have little effect on the outcome. This allows us to achieve a better tracking accuracy in handling partial occlusions and appearance changes. 2) Our algorithm can update the fragment template online during the tracking process. In the small fragments the method detects the possible fragments where appearance changes or occlusions occur, and then takes corresponding update strategies according to the matching of the small fragments. The fragment template overcomes the shortcoming of the traditional template, which modelled the whole target without the spatial information.

The rest of this paper is organized as follows. The subsequent section describes the fragment tracking algorithm. Then an adaptive multi-cue integration model is given in Section 3. Section 4 presents a new tracking algorithm of multi-cue integration based on fragment, which is the main contribution of this paper. Section 5 is devoted to algorithm analysis, including experimental results on videos with different tracking conditions, and gives qualitative and quantitative analysis. The final section concludes this paper and suggests the future direction of our research.

2 Fragment tracking algorithm

Fragment tracking is a kind of target tracking algorithm based on cues proposed by Adam [12]. The basic idea is dividing target window into multiple sub-fragments and represents the target according to the joint sub-fragment histograms. This kind of algorithm has good anti-interference and anti-occlusion, and the computation has nothing to do with tracked target size. Detailed description is as follows:

The division pattern is shown in Figure 3. The target template is:

$$T = \left\{ (h_{p_k}(b), \lambda_{p_k}) \right\}_{p_k=p_1, \dots, p_n}$$

p_k is the index of fragment k , the target is divided into n fragments, $h_{p_k}(b)$ and λ_{p_k} are a histogram and weight respectively, where $b=1, \dots, B$, and fragment weight equals 1 for every fragment.

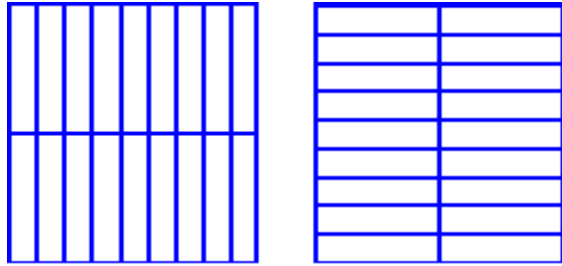


FIGURE 3 Vertical and horizontal fragment template

Fragment tracking is to find the most similar region to the template in current frame. The tracking principle is shown in Figure 4. Assume that a previous estimate of the position has been got, and each point in the neighbourhood of this estimate is searched, then every coordinate point in the searching window is a candidate for the target in current frame.

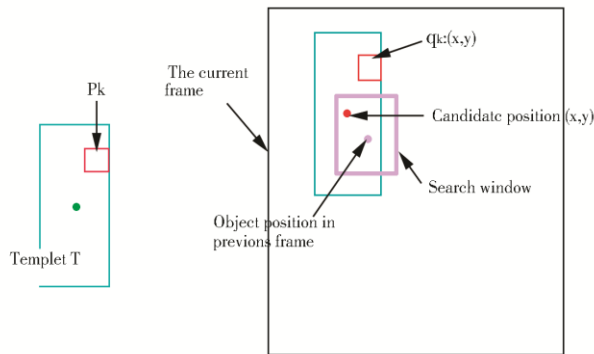


FIGURE 4 Fragment tracking principle

Let $p_k=(dx,dy,h,w)$ be a rectangular patch in the template, whose centre is displaced (dx,dy) from the template centre, and whose half width and height are w and h respectively. Let (x,y) be a hypothesized position of the object in the current frame. Then the patch p_k defines a corresponding rectangular patch in the image $q_{k:(x,y)}$ whose centre is at $(x+dx,y+dy)$ and whose half width and height are w and h , respectively. Calculating the histogram matching distances of each $q_{k:(x,y)}$ in the candidate region and the corresponding fragment in the template, we can get Formula (1) using the linear weighting of all the matching distances to represent the similarity between candidates and target template.

$$S(x, y) = \sum_{k=1}^n d(p_k, q_{k:(x,y)}) \lambda_{p_k}, \quad (1)$$

where $S(x, y)$ is the similarity between the candidate at the location (x, y) and the template, and $d(p_k, q_{k:(x,y)})$ is

the histogram matching distance between fragment p_k and fragment $q_{k:(x,y)}$, which is defined as

$$d(p_k, q_{k:(x,y)}) = \sum_{b=1}^B \frac{(h_{p_k}(b) - h_{q_{k:(x,y)}}(b))^2}{h_{p_k}(b) + h_{q_{k:(x,y)}}(b)}. \quad (2)$$

The smaller $d(p_k, q_{k:(x,y)})$ means more similarity between the histograms. After (x, y) has traversed all the candidate points, we obtain the similarity between each candidate target and the current template. Then the current target location (\hat{x}, \hat{y}) is defined as:

$$(\hat{x}, \hat{y}) = \arg \min_{(x,y) \in \Theta} (S(x, y)), \quad (3)$$

where Θ is the set of all the candidate locations. To distinguish easily between occlusions and appearance changes, after finding the target location in current frame according to the matching distance between fragment $q_{k:(\hat{x}, \hat{y})}$ in the target region and the corresponding fragment p_k , we update the weight of p_k as

$$\lambda_{p_k} = \exp\left(\frac{-d(p_k, q_{k:(\hat{x}, \hat{y})})}{\sigma_d^2}\right), \quad (4)$$

where σ_d is the variance of $d(p_k, q_{k:(\hat{x}, \hat{y})})$.

For convenience $q_{k:(\hat{x}, \hat{y})}$ is represented as q_k in the sequel.

3 Adaptive multi-cue integration

In complicated scene, it is difficult to obtain good tracking performance by single visual cue. To further improve tracking robustness, we need integrate other cues. Suppose F_j is a visual cue which has been selected and the fused cue is f_{fusion} , then f_{fusion} can be represented as:

$$\begin{cases} f_{fusion} = \sum_{j=1}^n \pi_j F_j \\ \sum_{j=1}^n \pi_j = 1 \end{cases}, \quad (5)$$

where π_j is the weight of F_j . It is clear that different cues have different contributions to the ability of identifying the whole target due to the impact of the environment, so different cues should have different weights. Since there may be target's appearance changes or occlusions in real tracking, the weight of each cue in

the each frame of video sequence should be updated dynamically.

There are two problems to be solved: one is how to evaluate the weights according to different contribution of each cue in different tracking environment, the other is how to update the weights in a steady and real-time way.

4 Proposed tracking algorithm

4.1 TRACKING PROCESS

This paper proposes a new method of fragments based tracking with multi-cue integration and template updating. The algorithm describes the target object by a template, and carries out tracking by searching for the image region in each frame of the tracking sequence with multiple cues similar to the template. This estimation process is done by splitting the target object into multiple arbitrary patches with each of them describing a different part of the target and accordingly providing the multi-cue information. The template is adaptively updated in accordance with whether occlusions or appearance changes occur to alleviate the template drift. The tracking schedule is shown in Figure 5.

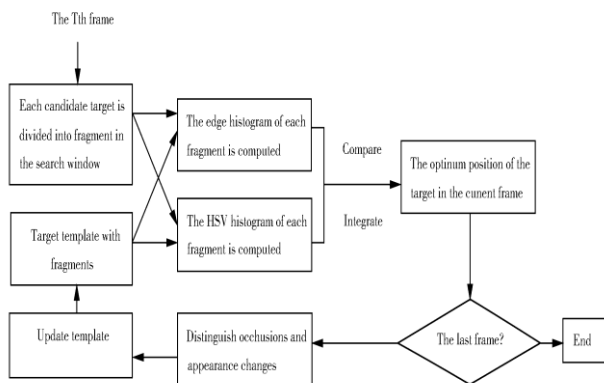


FIGURE 5 The flow chart of our algorithm

The steps of our algorithm are as follows.

Step1: Target template is divided into multiple fragments using the patch layouts given in Figure 2, and then each of these fragments is represented with histograms of different cues (colour, edge), and Integral Histogram data structure [18] is used to estimate them in a fast way.

Step2: The search window size is set to be $(2M + 1) \times (2M + 1)$ in current frame, namely the search radius is set to be M pixels from the previous target position (M =7 in this paper, and the search window is 15×15). Matching each template patch and the corresponding image patch is then carried out by comparing their histograms using the formula (13), and then we obtain the values $\{d(p_k, q_{k:(x,y)})\}$ by adaptive weight integration. Each $d(p_k, q_{k:(x,y)})$ indicates every fragment votes on the possible positions of the target. The detail is given in section 4.3.

Step3: The individual votes of the template patches are combined to obtain the joint tracking result. The weights of the template fragments are computed according to the formula (4), if $\lambda_{p_k} < 0.5$, the histograms of q_k changed a lot compared to corresponding p_k , that means there might be occlusions or appearance changes. We call such q_k as invalid fragment. The joint similarity $S(x, y)$ is defined in formula (6), and then we can get the optimum position of the target in current frame using formula (3):

$$S(x, y) = \sum_{k=1}^n \begin{cases} d(p_k, q_{k:(x,y)}) \lambda_{p_k}, & \lambda_{p_k} \geq 0.5, \\ 0, & \lambda_{p_k} < 0.5. \end{cases} \quad (6)$$

Step4: We judge each invalid fragment to see if it is caused by occlusion or appearance change. If there is occlusion, then go to step 2 directly without template update and start the next frame tracking; otherwise update the template and then go to step 2. The detail is described in section 4.4.

4.2 CUES HISTOGRAM COMPUTATION

In complicated scene, it is difficult to obtain good tracking performance by single visual cue. To improve tracking robustness, we need integrate other cues. It is very important to select the strong discriminate cues. This paper uses the integration of colour and edge cues to construct likelihood function for the purpose of scattering the impacts of all kinds of changes.

Colour is a major cue to describe the target. There have been many studies about the description of target colour cues before. Ref. [19] proposed a colour histogram to describe the target colour cues, which has simple calculation and fast processing, and is relatively robust in solving problems such as partial occlusion, rotation, etc. HSV histogram is employed to extract colour cue in each fragment in this paper. In our experiments, the Hue, Saturation and Value are quantified to be 16, 4, and 4. The three colour components are combined into one dimensional cue according to the quantitative level:

$$C = 16H + 4S + V, \quad (7)$$

where $C \in [0, 1, \dots, 255]$, so we can get 256 bins HSV histogram.

Edge is another commonly used cue to describe the target. This paper chooses gradient direction histogram to describe the edge cue of the target [19]. The calculation formulas are as follows:

Horizontal gradient

$$G_x(x, y) = f(x + 1, y) - f(x - 1, y), \quad (8)$$

Vertical gradient

$$G_y(x, y) = f(x, y + 1) - f(x, y - 1), \quad (9)$$

Gradient magnitude

$$m(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)}, \quad (10)$$

Gradient direction

$$\theta(x, y) = \arctan(G_y(x, y) / G_x(x, y)), \quad (11)$$

where $\theta(x, y) \in [-\pi / 2, \pi / 2]$.

$\theta(x, y)$ should be mapped into $0 \sim 2\pi$.

$$\sigma(x, y) = \begin{cases} \pi + \theta(x, y) & G_x < 0 \\ 2\pi + \theta(x, y) & G_x > 0 \text{ and } G_y < 0 \\ \theta(x, y) & G_x \geq 0 \text{ and } G_y \leq 0 \end{cases} \quad (12)$$

$[0, 2\pi)$ is divided into k bins with the width of each bin being $2\pi / k$ ($k=9$ in the paper). Each bin value is got by counting the pixels number in it and we can obtain the edge histogram.

4.3 ADAPTIVE WEIGHT INTEGRATION

In order to adapt to the environment and the target appearance changes, the weight of each feature in each frame must be updated in real time. Ref. [21] proposed Kolmogorov-Smirnov similarity measure criteria, K-S distance for short, to calculate the degree of similarity between the candidate object and target template. Its value represents the difference of two histograms; the smaller K-S value means smaller difference and higher similarity. Based on K-S distance, the matching distance for one cue is defined as follows:

$$d(p_k^j, q_{k(x,y)}^j) = \sum_{b=1}^B \frac{(h_{p_k^j}(b) - h_{q_{k(x,y)}^j}(b))^2}{h_{p_k^j}(b) + h_{q_{k(x,y)}^j}(b)} \quad (13)$$

$d(p_k^j, q_{k(x,y)}^j)$ is the histogram matching distance for j cue between fragment p_k and fragment $q_{k(x,y)}$. We can change the weight of corresponding cue by formula (14), so as to automatically change the influence of different cues to global functions. An adaptive weight integration formula is defined as in (15):

$$\omega_{q_{k(x,y)}^j}^j = \frac{d(p_k^j, q_{k(x,y)}^j)}{\sum_{j=1}^n d(p_k^j, q_{k(x,y)}^j)}, \quad (14)$$

$$d(p_k, q_{k(x,y)}) = \sum_{j=1}^n \omega_{q_{k(x,y)}^j}^j d(p_k^j, q_{k(x,y)}^j) \quad (15)$$

It can be seen that automatically changing the different cues weights can better reflect the contributions of cues to the result. In experiments $j=1, 2$, colour is the first cue, edge is the second, $d(p_k^1, q_{k(x,y)}^1)$ and $d(p_k^2, q_{k(x,y)}^2)$ are the histograms matching distances for colour and edge cues between fragment p_k and fragment $q_{k(x,y)}$ respectively. When the object begins to enter the occlusion or changes appearance, the edge will change greatly, the weight of $d(p_k^2, q_{k(x,y)}^2)$ is larger, the weight of $d(p_k^1, q_{k(x,y)}^1)$ is smaller, and then the contribution of edge cue to $d(p_k, q_{k(x,y)})$ is larger. Therefore, this adaptive integration method can better reflect actual situation during tracking.

4.4 TEMPLATE ONLINE UPDATE

In order to ensure real-time tracking effect, this paper tracks the target by dividing the target into fragments, judges states of fragments, and adjusts update strategy according to whether there is occlusion or appearance change. The distance between the fragment and the template will be increased because of occlusion or appearance change. The template does not have to be updated in the case of occlusion while the template needs to be updated in the case of appearance change.

Fragment weights of the template are given in Section 2. After the experiments we found that if $\lambda_{p_k} < 0.5$, then q_k is an invalid fragment. Through the analysis we found that if the invalid fragment is caused by occlusion, the H component in HSV space of the fragment will be distributed with greater probability in the previous frame in the background region; if it is caused by target's own changes, the H component in HSV space of the fragment will be distributed with greater probability in the previous frame in the target region. Because obstacles only appear around the target, we select a "ring" background region around the target to determine whether the target is occluded or not, where the area of the annular region is about 2 times of target area. The distribution of the H component in HSV space of the invalid fragment is back projected onto the target and annular background in the previous frame, so we can obtain the distribution region in the previous frame, and then calculate respectively the probabilities of the invalid fragment belonging to the target region and the background region. This paper uses the main H component in HSV space of the invalid

fragment for back projection, and the projection method is the same as the one used in Camshift [2]. As shown in Figure 6, back projection image is a binary image, where white pixels denote the pixels whose grey values are equal to those of the main H component of invalid fragments.

In back projection image, we count the number of the foreground pixels in the target area S_0 and the number of the foreground pixels in the whole image S_t . If the probability p_0 of the H component of the invalid fragment belonging to the target satisfies formula (16), we think the target appearance changes. th is set to 0.8 after many experiments. If the invalid fragment is determined as occlusion, the template will not be updated; if the invalid fragment is determined as appearance change, the information of the current frame will be updated to the template. So that we can rapidly get the target change information and suppress background disturbance to the template at the same time, which obviously improves the tracking performance:

$$p_0 = \log \frac{S_0}{S_t - S_0} > th \quad (16)$$



FIGURE 6 Back projection images under appearance change or the occlusion

5 Experiments and analysis

We use the following four groups of experiments to verify the effectiveness of the algorithm. The experimental environments are Microsoft visual studio 2008 and Opencv2.0. The results are analysed in Matlab 2008. Qualitative and quantitative analysis are made for the results of the four methods, the fragment-based tracker [12], the colour-texture based mean-shift tracker [7], the decentralized template tracker [17], and the method proposed in this paper (our method). In the figures, the window of our method is labelled as 1, the tracking window of the decentralized template tracker is labelled as 2, the tracking window of the colour-texture based mean-shift tracker is labelled as 3, and the tracking window of the fragment-based tracker is labelled as 4. The qualitative analysis is done through observing test images and the quantitative analysis is done through measuring location errors.

Let

$$err(x_{obj}, y_{obj}) = \sqrt{\frac{(x_{obj} - x_{true})^2 + (y_{obj} - y_{true})^2}{2}} \quad (17)$$

Then $err(x_{obj}, y_{obj})$ is the location error between the target location (x_{obj}, y_{obj}) and the true location (x_{true}, y_{true}) .

5.1 THE QUALITATIVE ANALYSIS

Test Sequence 1. Figure 7 shows the simulation results of a video in a campus named “campus”. The video sequence contains 500 384×288-pixel colour images. In the video, from frame 272 to frame 377, the target enters into dark region from bright region and later returns to the bright region. The illumination has a great change. At frame 347, the tracking window of the fragment-based tracker loses the target; the colour-texture based mean-shift tracker does not lose the target, but the estimation location of the target has already deviated from the actual location. At frame 387, the tracking window of the colour-texture based mean-shift tracker loses the target too while the tracking windows of our method and the decentralized template tracker never lose the target.

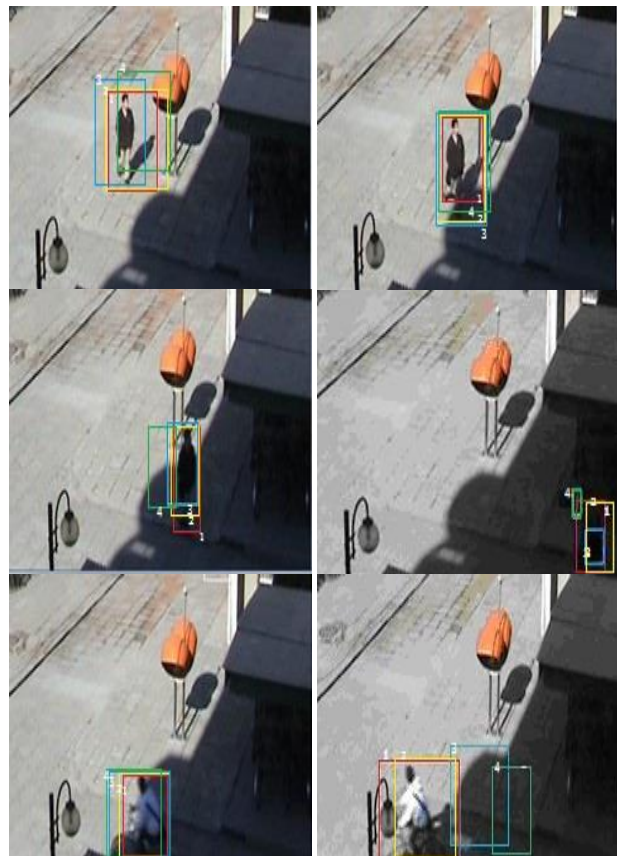


FIGURE 7 Tracking results of video “Campus” (Frames: 246, 272, 302, 347, 377, 387)

Test Sequence 2. Figure 8 shows the simulation results of video “hall”, which comes from CAVIAR database. The video sequence contains 300 384×288-pixel colour images. The main difficulty with this sequence is that the target is occluded by another people. When the target is occluded at the video frame 191, both the fragment-based tracker and the decentralized template

tracker lose the target from frame 208 to frame 221, and later on the colour-texture based mean-shift tracker loses the target too, while our method has kept tracking the target. To sum up, our method can always track the target no matter if there is occlusion or not.

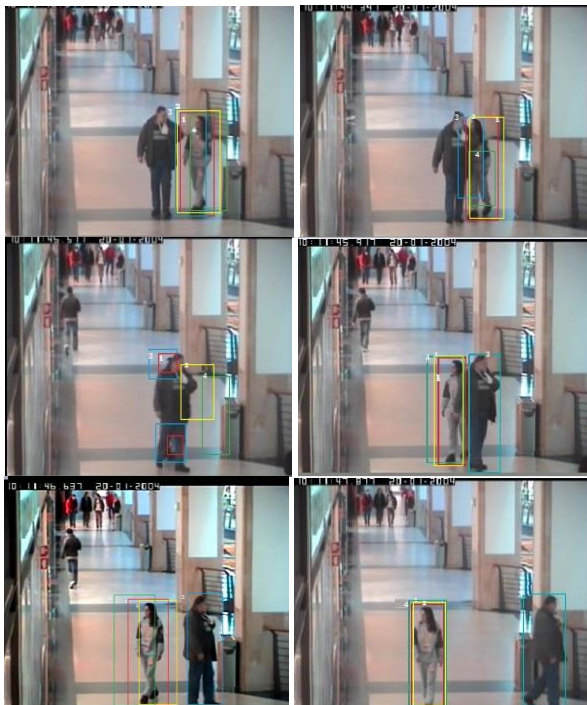


FIGURE 8 Tracking results of video “hall”
(Frames: 163, 191, 208, 221, 240, 267)

Test Sequence 3. Figure 9 shows the simulation results of video “man1” which comes from SPEVI database.

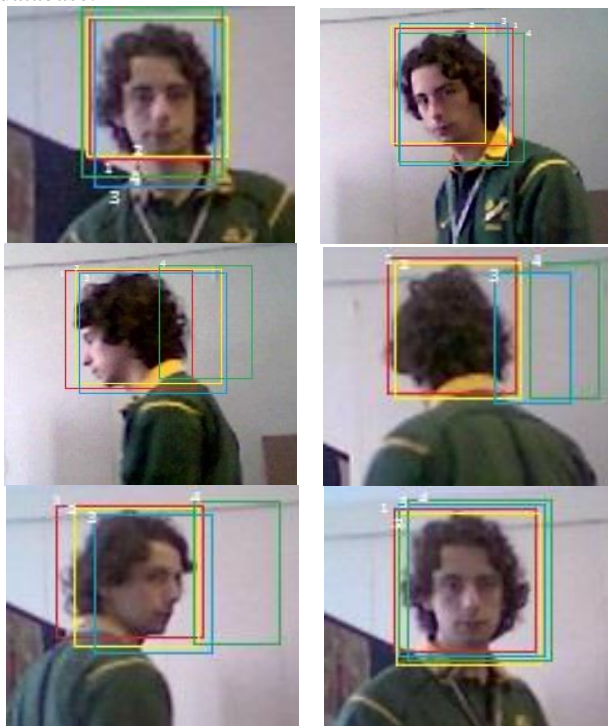


FIGURE 9 Tracking results of video “man1”
(Frames: 100, 110, 117, 130, 141, 145)

The video sequence contains 150 320×240-pixel colour images, where the man’s head in the sequence has irregular translation and rotation movements. At frame 100 and frame 110 when the face is approximately front face, the four methods track the target accurately. When the back face appears at frame 130, the fragment-based tracker and the colour-texture based mean-shift tracker lose the target immediately, and the fragment-based tracker cannot recover from the error even for a side face at frame 131. Our method and the decentralized template tracker have more robustness.

Test Sequence 4. Figure 10 shows the simulation results of video “man2” which comes from SPEVI database.

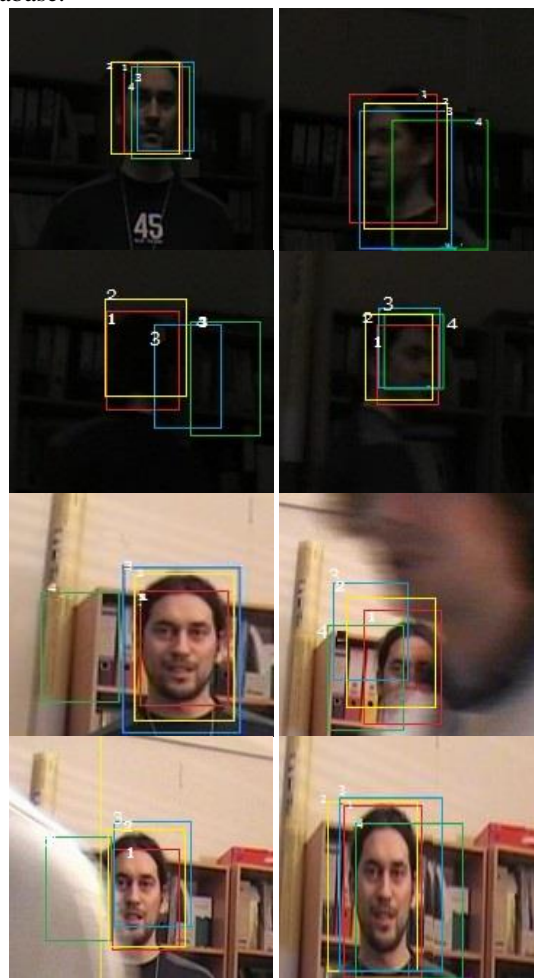


FIGURE 10 Tracking results of video “man2”
(Frames: 368, 384, 398, 410, 474, 580, 590, 601)

The video sequence contains 600 320×240-pixel colour images. The target first rotates and then the illumination changes sharply, followed by the part occlusion in the sequence. At frame 398 when the target rotates, both the colour and gradient of the target change, and the fragment-based tracker with one single cue tracking loses the target when the target turns back, the colour-texture based mean-shift tracker with fixed template tracking loses the target too, and the decentralized template tracker brings about a certain tracking drift while our method achieves accurate

tracking. At frame 474, the lamp is turned on, which makes the illumination change sharply and the fragment-based tracker lose the target again, while the other three algorithms keep tracking the target. At frame 580 when the target is occluded by another moving object, the other three algorithms fail to track the target accurately while our method still tracks well, since our method not only uses adaptive cue integration scheme during fragment

tracking, but also employs the online updating template by distinguishing appearance changes and occlusions.

5.2 THE QUANTITATIVE ANALYSIS

Figures 7-10 show the qualitative analysis results. We also provide the error plots for each sequence in Figure 11.

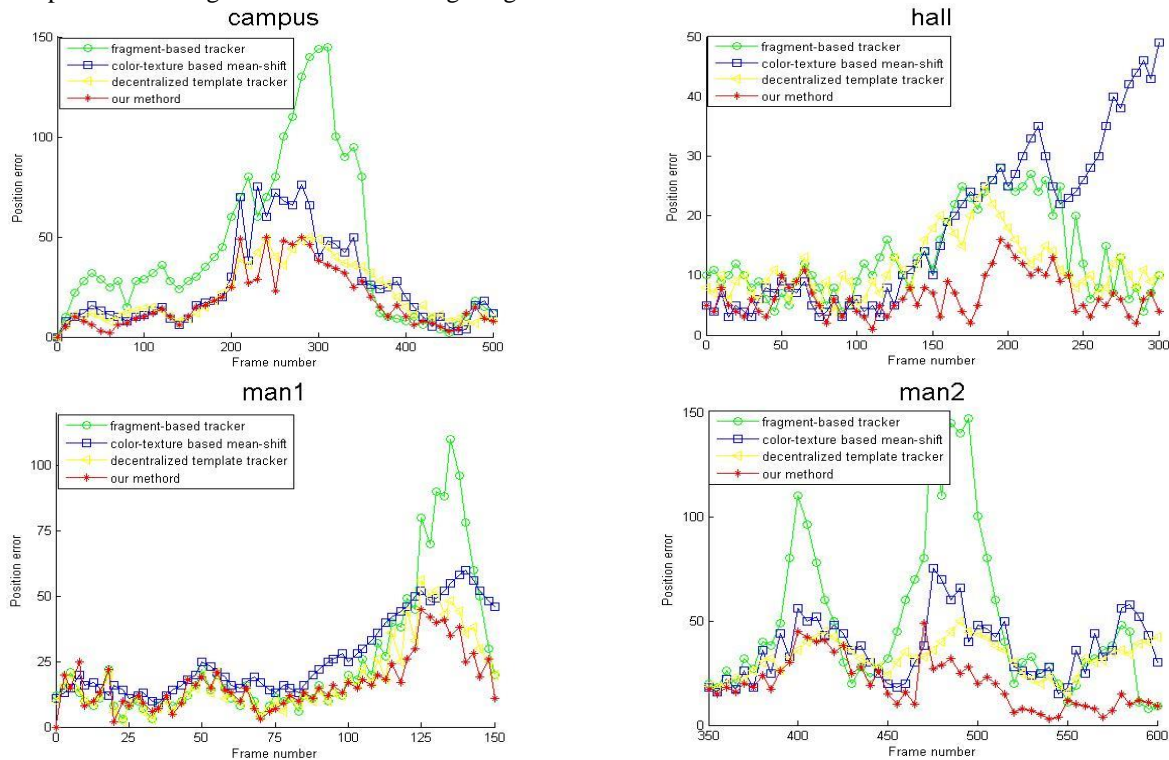


FIGURE 11 Error plots for the video sequences used in the quantitative analysis

It can be seen from these results that the proposed algorithm outperforms the fragment-based tracker in terms of tracking accuracy. The reason for this mainly stems from our adaptive cue integration scheme, which removes the assumption on the degree of potential occlusions for the fragment-based tracker, and replaces its competitive approach to cue integration with a more cooperative strategy.

The colour-texture based mean-shift tracker, which uses the joint colour and LBP cues to track target, outperforms the single cue tracking approaches and is robust to the complex environment. However, it yields, in many cases, poor tracking results when the target becomes occluded by the other objects in the scene. Our method and the fragment-based tracker cope with these kinds of occlusion better than most of the multi-cue integration trackers. The reason stems from tracking the target in fragments.

The decentralized template tracker, which uses online updating template to track target, is proved to be very robust against severe appearance changes since it can learn a new object model at each frame. Our method gives better results than the decentralized template tracker

for the video sequences because of the fact that the decentralized template tracker updates target template at each frame without discrimination of appearance changes and occlusions, and cannot prevent the template from drifting, even if it uses short-term and long-term appearance models

As a result, for most of the video sequences, our algorithm provides the best results.

6 Conclusions

Video target tracking algorithm is faced with many problems, such as illumination change, occlusions, appearance changes, etc. A new tracking algorithm of multi-cue integration based on fragment has been proposed in this paper. The new algorithm associates the image fragments describing the different parts of the target object with multi-cue integration, where the weight of each cue is dynamically adjusted during tracking. In this way, the vote of each fragment contributes to the joint tracking result according to its adaptive weight integration. Furthermore, the algorithm uses the template updating for target tracking, designs the corresponding




template update algorithm, and successfully implements the target tracking under complex background. We have demonstrated the potential and the effectiveness of the proposed approach on various challenging video sequences with different tracking scenarios. As revealed by our experiments, the proposed approach works generally better than the fragment-based tracker, the multi-cue integration trackers and template-based trackers for the sequences in the complex background of illumination, occlusion, and the appearance change.

Acknowledgements

The authors wish to thank the helpful comments and suggestions from my teachers and colleagues. This work was supported by the National Natural Science Foundation of china(Grant No.31371533), National Science and Technology Support Program(Grant No. 2013BAJ10B12), the Key Technologies R & D Program of Anhui Province (Grant No.1301032169) and by the Natural Science Foundation of Anhui Province(Grant No.1308085MF89).

References

- [1] Birchfield S 1998 *Computer Vision and Pattern Recognition (CVPR) Santa Barbara, CA:IEEE Computer Society* 232-7
- [2] Triesch J, von der Malsburg C 2001 *Neural Computation* **13**(9) 2049-74
- [3] Yuru Wang, Xianglong Tang, Qing Cui 2012 *Pattern Recognition* **45**(12) 4510-24
- [4] Brasnett P, Mihaylova L, Bull D, Canagarajah N 2007 *Image Vision Computation* **25**(8) 1217-27
- [5] Wang J Q, Yagi Y 2008 *IEEE Transaction on Image Processing* **17**(2) 235-40
- [6] Jia G, Tian Y, Wang Y 2010 Dynamic multi-cue tracking with detection responses association *Proceedings of the ACM Multimedia NY USA: ACM New York* 1171-74
- [7] Ning J, Zhang L, Zhang D, Wu C 2009 *International Journal of Pattern Recognition and Artificial Intelligence* **23**(7), 1245-63
- [8] Li Ying, Liang Shi, Bai Bendu, Feng David 2012 *Multimedia Tools and Applications* **6**(3) 1-21
- [9] Nejhum S M S, Ho J, Yang M H 2008 Visual tracking with histograms and articulating blocks *Proc. Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage, AK :CVPR 2008* 1-8
- [10] Nickel K, Stiefelhagen R 2008 Dynamic integration of generalized cues for person tracking *Proc. Eur. Conf. Computer Vision (France) (ECCV): Marseille* 514-26
- [11] Chockalingam P, Pradeep N, Birchfield S 2009 Adaptive fragments-based tracking of non-rigid objects using level sets *Proc. Int. Conf. Computer Vision (Japan) (ICCV) :Kyoto* 1530-7
- [12] Adam A, Rivlin E, Shimshoni I 2006 Robust fragments-based tracking using the integral histogram *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)* 798-805
- [13] Erdem E, Dubuisson S 2012 *Computer Vision and Image Understanding* **116**(7) 827-41
- [14] Porikli F, Tuzel O, Meer P 2006 Covariance tracking using model update based on Lie algebra. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition:New York, USA* 728-35
- [15] Collins R T, Liu Y X, Leordeanu M 2005 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(10) 1631-43
- [16] Wang Yong-Zhong, Liang Yan, Zhao Chun-Hui, Pan Quan 2008 *Acta Automatica Sinica* **34** (4) 393-9
- [17] Firouzi Hadi, Najjaran Homayoun 2012 *Pattern Recognition* **45**(12) 4494-509
- [18] Porkili F 2005 Integral histogram: A fast way to extract histograms in Cartesian spaces *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* 439-47
- [19] Comaniciu D, Ramesh V, Meer P 2003 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(5) 564-77
- [20] Yang C, Duraiswami R, Davis L 2005 Fast multiple target tracking via a hierarchical particle filter *Proceedings of the Tenth IEEE International Conference on Computer Vision: Beijing, China* 212-9
- [21] Simard R, L'Ecyer P 2011 *Journal of Statistical Software* **39**(11) 1-18
- [22] Fashing M, Tomasi C 2005 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(3) 471-4

Authors	
	<p>Lichuan Gu, October 1974, Sichuan, China</p> <p>Current position, grades: Associate Professor of Computer Science Department of Anhui Agricultural University University studies: BSc, MSc degrees in computer science in HeFei University of technology, China, in 1997 and 2005. Scientific interest: research interests, mainly including machine learning, image processing, and pattern recognition.</p>
	<p>Chengji Wang</p> <p>Current position, grades: master studies in agricultural informatization in AHAU University University studies: BE in Logistics Engineering, AHAU University (2013) Scientific interests: research interests include cognitive computing and logistics system optimization</p>
	<p>Jianxiao Liu</p> <p>Current position, grades: master studies in Agricultural Informatics at University. University studies: BE in Logistics Engineering (2013) from AHAU University. Scientific interests: research direction includes different aspects of Agricultural logistics</p>

FPGA based accelerator for parallel DBSCAN algorithm

Shaobo Shi, Qi Yue, Qin Wang*

School of Computer and communication engineering, University of Science and Technology Beijing

Received 1 January 2014, www.tsi.lv

Abstract

Data mining is playing a vital role in various application fields. One important issue in data mining is clustering, which is a process of grouping data with high similarity. Density-based clustering is an effective method that can find clusters in arbitrary shapes in feature space, and DBSCAN (Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise) is a basic one. With the tremendous increase of data sizes, the processing time taken by clustering algorithms can be several hours or more. In recent years, FPGA has provided a notable accelerating performance in data mining applications. In this paper, we study parallel DBSCAN algorithm and map it to FPGA based on the task-level and data-level parallelism architecture. Experimental results show that this accelerator can provide up to 86x speedup over a software implementation on general-purpose processor and 2.9x over a software implementation on graphic processor.

Keywords: Data mining, Clustering, Parallel DBSCAN, FPGA, Hardware Accelerator

1 Introduction

Clustering is an effective approach of retrieving useful patterns from raw data sets. The process of clustering is to group data into different clusters so that objects in the same cluster have high similarity. And clustering is an important data mining tool that has applied in many areas. DBSCAN is an effective density-based clustering method, which is proposed by Martin [1]. Compared with other clustering algorithms, DBSCAN has some obvious advantages such as requiring minimal domain knowledge, being able to discover clusters in arbitrary shapes, being robust in removing noise and outliers, and having a good efficiency on large databases.

However, performing DBSCAN algorithm in practice is limited by the fact that the performance of general processors is improving at slower rate comparing to rapidly growing data set size. For this reason, a preferable method of accelerating executing speed is to make an algorithm parallelized. Some researchers proposed parallel DBSCAN algorithm and mapped it to distributed parallel computing platform, and achieve a near-linear speedup performance [2, 5]. However, this solution is not energy efficient since distributed parallel computing platform requires high power consumption.

Field-programmable gate arrays (FPGAs) are used as user-customized computing engines for accelerating a wide range of applications. The high-end FPGAs are characterized with enormous amount of logic gates, abundant on-chip memory and large capacity external storage, flexible programmability and lower power consumption. With these features, users can utilize

multiple operation-levels and high memory access bandwidth for specific applications. Thus, we take FPGA as hardware platform and study the mapping from parallel DBSCAN algorithms to FPGA to get a higher speedup performance in this paper.

Mapping a parallel DBSCAN algorithm to FPGA should consider two issues. The first one is how to fully exploit parallel DBSCAN algorithm with the high flexibility of FPGA structure. In addition, the second is how to eliminate data dependencies existing in parallel algorithm. The contributions of this paper are summarized as follows:

- ✓ We propose a hardware architecture based on task-level and data-level parallelism, which fully exploit the bit-level parallelism provided by FPGA.
- ✓ We design a data reused pipeline structure to eliminate the extra memory access caused by the data dependencies in parallel algorithm.
- ✓ Based on previous work, we propose a more robust parallel algorithm. which can avoid wrong clustering results in some special conditions.
- ✓ To the best of our knowledge, it is the first work of implementing a parallel DBSCAN algorithm on reconfigurable hardware. Compared to the sequential software implementation on Intel general processor, our accelerator can achieve 80x speedup. Besides, we can get a 2.9x speedup over the similar parallel implementation on Nvidia graphic processor.

The rest of this paper is organized as follows. In section 2, we review the previous work on parallel DBSCAN implementations and hardware accelerations for the data mining applications. We introduce the

* Corresponding author - E-mail: 337816437@qq.com

original DBSCAN algorithm and propose complete robust parallel DBSCAN algorithm in section 3. In section 4, the accelerator hardware architecture and the specific design are presented. The performance model and experimental result are discussed in section 5. Section 6 is conclusion.

2 Related Work

Parallel DBSCAN implementation: Several works for parallelizing DBSCAN algorithm are proposed. PDBSCAN [2], whose parallel partitions are based on regionalism, is an implementation on the master-slave mode computer cluster [3] maps the sequential kernel of DBSCAN to a higher level parallel programming environment [4] is another implementation focus on parallelization of DBSCAN through simple distance function. In [5], author presented an improved work to [1] that mapped parallel algorithm to the MapReduce framework. All of these researches are based on the region partitions, and use shared-nothing architecture to run parallel algorithm. Meanwhile, the parallelism of these methods is too low, so they are not suitable for being implemented on FPGA.

CUDA-DClust is implemented with GPU [6], which is a fine-grained parallel DBSCAN algorithm and it is different from the above mentioned works. Moreover, this work can avoid the boundary processing and load unbalance issues caused by region partitions. Due to these advantages, CUDA-DClust is a good reference and comparable object for our work. However, due to the long communication delay between different Streaming Multiprocessors in GPU, the data dependency in parallel DBSCAN algorithm will increase the memory access times and thus influence the performance when it mapped on the GPU. The experiments in [7, 8] point out a fact that implementation of a customized data path in FPGA can provide a superior performance over GPU in the presence of data dependency. Therefore, the optimized and customized hardware architecture for parallel DBSCAN appears very necessary. Compared with the software implementation of DBSCAN in parallel computers system and GPU, our approach is far better than these methods in both performance and power consumption.

Hardware accelerator for data mining applications: There have been many prior researches on hardware implementation of data mining algorithms. In [9], K-means clustering is implemented as a reconfigurable accelerator, which simplified the distance calculation. However, K-means is essentially different from DBSCAN since it is a partitioned algorithm. In [10], the kernel of HOP algorithm is implemented on FPGA platform. Although HOP is a kind of density-based clustering algorithm, the final goal of this algorithm is to find the nearest densest neighbours rather than the transitive closure computation in DBSCAN. In addition,

those works did not work on the parallelism from the view point of algorithm. In [11], hardware architecture for Decision Tree Classification (DTC) algorithm is described.

Other hardware implemented clustering algorithms are summarized as follows. The Apriori algorithm, a popular association rule mining algorithm, is accelerated by systolic array architecture in [12]. And its improved work with bitmapped CAM is proposed in [13]. The HAPPI architecture is proposed in [14] with the pipeline and hashing methodology to resolve the bottleneck of Apriori. In [15] the FP-Growth algorithm is firstly mapped to a systolic tree structure by mimicking the internal memory layout of software algorithm.

3 Sequential DBSCAN algorithm and Parallel model

3.1 SEQUENTIAL DBSCAN ALGORITHM

The original DBSCAN algorithm is a sequential clustering algorithm [1]. The key idea of DBSCAN is that the data density within a small area in the feature space must exceed a given threshold, i.e., the neighbourhood of each point in a cluster must contain a minimum number of points. To make a clear presentation of the proposed method in this paper, we introduce some basic definitions of DBSCAN as follows:

1. $N_{Eps}(p)$ is a set: $\{q \in D \mid dist(p, q) \leq Eps\}$
2. *Directly density-reachable*: A point p is directly density-reachable from a point q in the set of point D if $p \in N_{Eps(q)}$ and $Numb(N_{Eps(q)}) \geq MinPts$
3. *Density-reachable*: A point p is density-reachable from a point q in the set of point D if there is a chain of points $p_1, \dots, p_n, p_1 = q, p_n = p$ such that $p_i \in D$ and p_{i+1} is directly density-reachable from p_i
4. *Density-connected*: A point p is density-connected to a point $q \in D$ such that both p and q are density-reachable from o w.r.t Eps and $MinPts$ in D
5. *Cluster*: A cluster C is a non-empty subset of D iff:
 - Maximality*: $\forall p, q \in D$: if $p \in C$ and $q \in N_{Eps(p)}$ then also $q \in C$.
 - Connectivity*: $\forall p, q \in C$: p is density-connected to q in D .
6. *Core point*: o is a Core point if $N_{Eps(o)} \geq MinPts$.
7. *Border point*: p is a Border point if $N_{Eps(p)} \leq MinPts$ and p is Directly density-reachable from a Core point.
8. *Nosie point*: p is a Nosie point if $N_{Eps(p)} \leq MinPts$ and p is not Directly density-reachable from any Core point.

With these definitions, we show the simple sequential proceeds of algorithm as follows:

- S1: For arbitrary unclassified point $P \in D$,
- S2: Retrieve(P), If P is core point, mark with ClusterID;
- S3: For the unclassified or noise point $Q \in N_{Eps}(P)$, store them into SeedStack, mark with ClusterID;

- S4: While SeedStack not empty, for the top element
P_top of SeedStack
- S5: Repeat S2-S4
- S6: If D have unclassified point, goto S2

The main task of DBSCAN is retrieve function (step S2) for find the neighbourhood of each point and finds the transitive closure relationship. The complexity of DBSCAN is $O(N^2)$ without index structure, and is $O(N \log(N))$ if with a multidimensional index structure. We run the sequential non-index algorithm in Intel Vtune Performance Analyzer Tool, the result shows that CPI rate is 0.761 and the Data Reference per instruction retired (Load/Store) is 0.742, according to the MineBench in [16]; DBSCAN is compute-intensive and memory-intensive. Our hardware architecture is designed for non-index DBSCAN, but it can be transplanted to fit the DBSCAN with index structure easily.

3.2 PARALLEL MODEL BASED ON THE TempCluster

From sequential algorithm, DBSCAN can start at an arbitrary point, which is unclassified to find the other points that fulfil the maximum density connectivity. Intuitively, we can unroll this loop so that DBSCAN can start at multiple unclassified points. This is the basic parallel concept that different clusters with respective ClusterIDs can be clustered simultaneously. These clusters are identical with the concept of ‘Chain’ proposed in [6]. Compared to the definition of cluster, a chain is a set of data object belong to a common density-based cluster that do not have to meet maximalist. In other words, a chain can be considered as a tentative cluster with a tentative clusterID. Different chains may have collisions that mean these chains belong to the same cluster, so the collision check mechanism is necessary.

Our hardware parallel model is based on chain, but we have a robust collision check mechanism that is described later. In order to distinguish ‘Chain’ from each other, we call it as TempCluster (TC). The definition of TempCluster is:

TempCluster is a subset $C \subseteq D$, if and only if :
 $\forall p, q \in C : p$ is Density - connected from q each other.

$$\text{Clusters} \subseteq D : \{C_1 \cup C_2 \cup C_3 \cup \dots \cup C_n\},$$

$$C_i \subseteq D : \{TC_{i1} \cup TC_{i2} \cup TC_{i3} \cup \dots \cup TC_{in}\}$$

Thus, the clustering results of D are composed of one or several C_i and each C_i is made up of one or several TC_{ij} . Therefore, we can devise the customized data path for the parallel generation of multiple TC_{ij} with different start points.

In [6], the valid condition of collision is the point in an arbitrary TC should be marked already has a different

TempClusterID (TCID). This means that this collision point has been marked by other TCs (step S2), so this point should not be store into SeedStack of local TC (step S3). However, the collision check criteria proposed in [6] may generate wrong clustering results under special condition. As shown in Figure1(a), the black point is this special collision point. This point is not a core point for its neighbour points are less than Minpts. TC1 and TC2 should not be merged into one cluster, because they cannot satisfy the density connected condition according to definition4, and. If the TC1 and TC2 are merged into one cluster, the clustering effectiveness of DBSCAN will be impacted seriously. We propose a robust collision check criterion as follows:

Let TempCluster1 and TempCluster2 be found in D , TempCluster1 \cup TempCluster2 belong to same cluster if and only if :
 $\exists o \in D, o \in TC1 \cap TC2$, and o
 1, o is border point in TC1, and o is core point in TC2;
 2, o is border point in TC2, and o is core point in TC1;
 3, o is core point in both TC1 and TC2.

This check method utilizes the concept of border property and core property, and can exclude the special situation that two clusters just share with a border point. As shown in fiugre1(b), the black point is a valid collision point. Therefore, our collision check criteria can avoid occurrence of special phenomenon described above. In our later implementation, the collision check procedure must check the border value and core value in each point data. A detailed description of our check method is presented in algorithm2 of Section4.

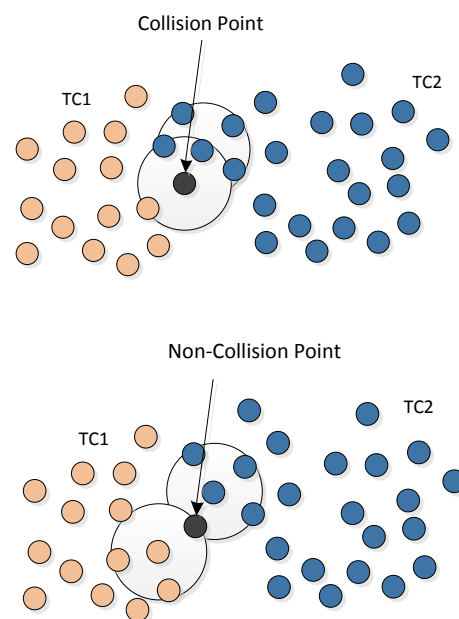


FIGURE 1 a) Invalid Collision, b) Valid Collision

4 Hardware architecture

As Figure 2 show, DBSCAN algorithm accelerator consists of FPGA chip(XC6VLX240T) and external memory which includes DDR3 SDRAM. The accelerator loads data from host PC to SDRAM through PCIe in DMA mode, and user can read the clustered data when the algorithm done from accelerator in the same way. We mapped parallel hardware architecture to FPGA. Some key module such as Processing Elements Array (PE Array) is responsible for perform the parallel TempCluster expansion. PE Array control unit is used as manages the control signal, and update the candidate data in PE Array.

And other modules such as memory interface (PCI-E interface, DDR3 SDRAM interface) and I/O FIFO are responsible for memory access and data buffer.

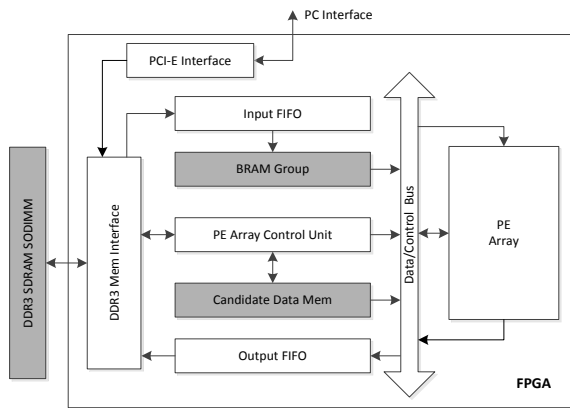


FIGURE 2 Architecture of Parallel DBSCAN accelerator

The raw data store in the SDRAM we call it Spatial Data (SPData), for example, one 2-dimision data occupy 64bits and is composed of several properties data. The format of this data is shown in the figure 3. The ID data has 20 bits that can support the representation of 10^6 data amount. Core point and neighbour point properties use 1 bit respectively, and ClusterID use 9 bits. The stack property also use 1 bit. We also call the high-order 32bits data as data property. Each coordinate value is a 16 bit fixed point data. For the high dimension data, we only need expend the data property.

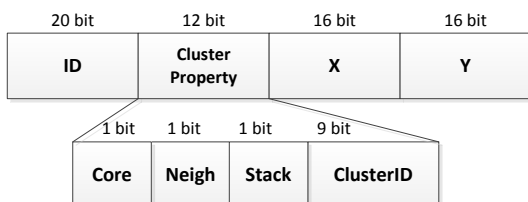
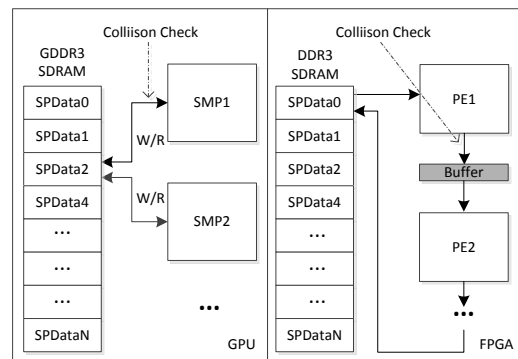


FIGURE 3 Data format in SDRAM

4.1 OPTIMIZED PE ARRAY DESIGN STRATEGY FOR COLLISION CHECK

We adopt multiple PEs to perform multiple TempCluster calculation in parallel pattern, and single PE is just responsible for one TempCluster. In GPU architecture, it is a similar processing method for parallel algorithm. However, data dependencies caused by collision check can decrease the performance when parallel DBSCAN executes on GPU. Left part of figure 4 demonstrates how data dependency affects GPU parallel implementation. In GPU, all of SMPs (StreamingMultiProcessors) share data through the Global Memory (GDDR SDRAM), and the SPData are stored in the Global Memory. Each SMP is responsible for a TC. For a simple example, SPData2 is read by SMP1, and meet the ClusterID marking condition through calculation. Thus the SPData2 should return its latest ID property to SMP1 in order to check whether it has been marked by other SMPs. And in this procedure is performed by an atomic operation provided by GPU software, other SMPs cannot access SPData2 before read operation of SMP1 complete. Since the situation that all SMPs have a large latency to Global Memory and memory access conflicts caused by data dependency, parallel DBSCAN algorithm is not so efficient when mapped to GPU architecture.

Aimed at this problem, a solution in hardware design that minimizes the overhead of data dependency is to make collision check between PEs in local memory and avoid memory access conflicts. And the abundant register resource and design flexibility of FPGA make this solution possible. As right part of Figure4 shows, the latest ID property SPData0 can be transmitted from PE1 to PE2 directly through the buffer. If SPData0 in PE2 should be checked, we can obtain its latest state from buffer without reading SDRAM. And this buffer can only be written by PE1 and read by PE2, so we can avoid memory access conflicts. The more detailed operation-level design will be described later.



Collision Check flow Comparison between GPU and FPGA Arch

FIGURE 4 Difference between GPU and FPGA in Memory allocation

4.2 PE ARRAY DESIGN BASED ON DATA REUSES AND TASK-LEVEL PARALLELISM

We propose a pipeline PE array based on the data reuse, to eliminate the extra memory access caused by collision check. As the Figure 5 show, This PE array can work in a fully task-level parallel mode. In the seed expansion, PE will calculate distance with every SPData in Dataset (with the index structure, is the all of the region data that

fulfil the query condition). Thus, this is make the data reuse possible. We design register files in each PE, called SpatiData Reg. And the SPData from SDRAM will the transfer to every PE cycle by cycle in the form of data streaming. PE#1 is responsible for the load data from SDRAM, other PEs load data from previous PE's SpatiData Reg. Lastly, PE#N will write clustered data back to SDRAM.

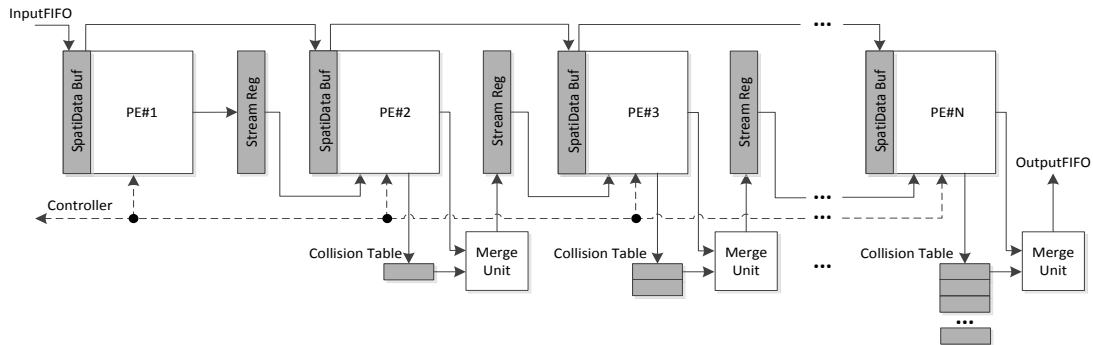


FIGURE 5 Architecture of PE array

The output of PE is the property data which has been compared with Eps, and its high-order 32-bits in SPData. We insert register files between each PE to buffer these output data, called Stream Reg. For the fact that PE#i+1 always fall behind PE#i one cycle, so buffering output one cycle can play an synchronous role when each PE perform collision check. The LableCheck Unit of PE#i+1 can get the output of PE#i simultaneously, and make collision check according to the criteria mentioned before. The algorithm executed on the PE Array showed in Algorithm 1.

Algorithm 1. Parallel Algorithm for PE Array

Parallel DBSCAN Algorithm for PE array in one Candidate corepoint Calc:
 Input: SpatialDataset D,e,Minpts
 Output: Clusterd SpatialDataSet ,CollisionSignal

```

-----
Initi: Generate Seed[N], TCID[N]
Dirtribute Dataset to PE#i one by one
For i=1:N PE#i Parallely Do:
    ExpandTempCluster(Seed[i],TCID[i],e,Minpts,Dataset,ClusterDataset);
    receive clusterd Dataset from Pre StreamReg;
    send clusterd Dataset to Next StreamReg;
    If PE#i return false
        Generate new seed to PEi;
        mark candidate point as Nosie;
    Else PE#i has Collision PE# X
        send TCID[i] and TCID[X] to Controller unit
        Update Collision Matrix;
    End IF
End For
    
```

This collision check method make PE obtained the latest data, and should generate collision signal without pipeline break off. From this figure we can see that the PE2 to PE N are likely to generate the collision signal,

and the generation is simultaneous. And the i-th PE has a potential collision signal with anyone among the first PE and (i-1)-th PE. Therefore, each PE must have a structure to record the collision signal. Due to the different amount of potential collision signal in each PE and the small storage amount, we adopt the register file to implement the function of collision signal record, we called it Collision Table. Besides the collision signal record unit, we also need a unit to process these collision signals, and it is called Merge Unit. The function of Merge Unit is to update the cluster property of every point with the new ClusterID according to the collision table.

4.3 PE DESIGN BASED ON DATA-LEVEL PARALLELISM

Since the high bandwidth of SDRAM, each PE can obtain several SPData in one clock, and PE can perform data-level computation. For the data which arrive at successive clock have no data dependency, so we can map sequential process procedure in pipeline. The DCU Array performs parallel Euclidean distance calculation between seed data and SPData, the distance formula is:

$$D_i = \sum_{j=1}^K (CD_j - SD_j)^2 \quad , \quad \text{where } CD = \text{CandiData}, SD = \text{SPData}, K - \text{dimension}.$$

The Figure 6 shows the PE structure. The parallelism degree depends on the data width and the SDRAM bandwidth. DCU array are composed with k subtractions (adder logic), k multipliers and one adder to get a distance.

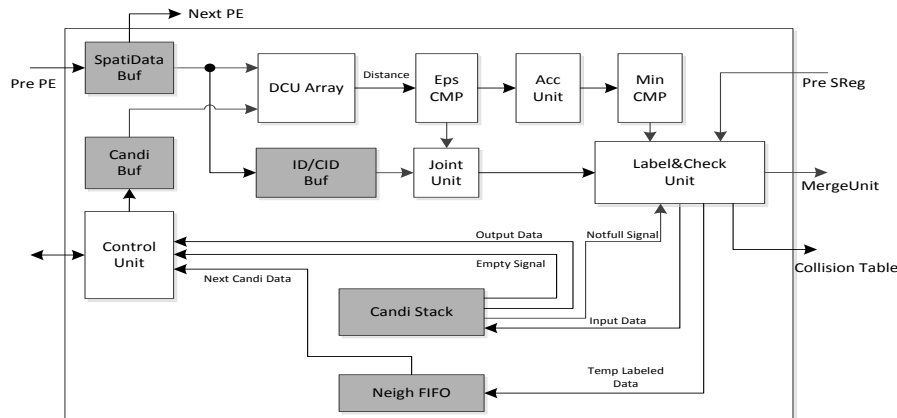


FIGURE 6 Architecture of PE

This is a three stages pipeline structure that can produce distance every cycle, and the multiplier is used IP core provided by Xilinx. For other module, Eps comp module, Acc Unit, MinComp, LabelCheck Unit are respectively complete the function of S22 to S23 in Algorithm 2.

Algorithm 2. PE Array Expand Algorithm in each PE

```

Parallel ExpandChain in each PE:
Input: Seed ,PointSet ,ε ,Minpts
Output: Clusterd PointSet ,CollisionSignal
-----
ExpandKernel(Seed,PointSet,TCID)
Initi phase:
S1:P=PointSet[[]];NeighborCount =0;
Calc phase:
S2:For i=0 to sizeof(PointSet) step DCUArraySize Do:
  S21:DCUArray parallelly Do:
    Q=PointSet[i];
    d=Distance[seed][Q];
  S22:EpsCMP and AccUnit parallelly Do:
    if d ≤ ε then
      h=atomicAcc(NeighborCount);
    end if
  S23:if h ≥ Minpts then
    if(Stack not full)
      store the ID of Q to CandiStack;
    else
      mark the stackbit of point with 1;
    end if
    CheckandMark(Q,TCID);
  else
    store the ID of Q to NeighborStack;
  end if
end for
S24:if h ≥ Minpts then
  mark seed as CorePoint;
else
  mark seed as Noise;
S3:while(CandiStack not empty) do:
  newseed = pop(CandiStack);
  send newseed to controllel unit;
  repeat S1-S2;
end while
  
```

ID/CID Buf is to buffer the property data before put them into LabelCheck Unit in pipeline. Candi Stack store the effective point, but limited to the capacity of BRAM on FPGA, the stack bit in overflow data must be marked as 1 to indicate this point is belong to these PE, and will be write to the DDR3 SDRAM. The detail executed process of PE is shown in Algorithm 2.

5 Experiment and Evaluation

5.1 PERFORMANCE MODEL

The total run time is composed of data transfer between host pc and accelerator, clustering time on FPGA and the time of merging TC into complete cluster through collision signal. And run time of clustering on FPGA is the most consuming time part.

Performance Model:

- Number of PE: p
- Frequency of FPGA: f MHz
- Data amount: M points
- Data width per each point: w Byte
- Input Bandwidth for each PE: $bandwidth$ Byte/s
- Pipeline Depth of PE: D

$$t_{actual_comp} = (\text{floor}(\frac{M}{p}) * (\frac{2 * m * w}{bandwidth} + D + 2 * P)) * \frac{1}{f}$$

This is an ideal model, P is the number of PE, and M is the amount of SPData. Data width for each SPData is w, and B is input bandwidth of PE#1. And f is frequency of FPGA. D is the depth in single PE. The DBSCAN hardware architecture was implemented on a Xilinx ML605 board, which includes a Virtex-6 LX240T FPGA, a 512MB DDR3 SDRAM. The peak bandwidth of SDRAM can achieve 6.4GB/s.

5.2 COMPARISON WITH PC

In order to evaluate the DBSCAN accelerator performance, we run sequential DBSCAN algorithm on Intel Corei7 920 quad-core CPU, 2.7GHz and 4GB RAM. The test data set is from SpatialDataGenerator[17]. We generate 10 groups 2 dimension data set from 100k to 1000k points, each data set contains 10 clusters, Minpts is 4, Eps is 0.02. The software implementation is written in C. For the simple control of SDRAM access, frequency

of FPGA is set to 200MHz (maxim speed shows in Table1).

5.3 COMPARISON WITH GPU

TABLE 1 Resource utilization

Data Dim	Number of PE	Slices	BRAM	Frequency
2	40	32445(85%)	14400Kb(96%)	260.586MHz
8	20	30300(80%)	14400Kb(96%)	197.570MHz

Thus, PE#1 can receive 256bits in one cycle since single 2-dimension SPData is 64bits. LX240T can support 40 PEs for 2-dimension SPData as Table1 show. Runtime comparison as Figure7 show and the average speed up over CoreI7 is 86x to 72x with 40 PEs by our accelerator.

Although the bandwidth of GDDR3 is ten times of DDR3, and the frequency of GPU is higher than FPGA, we also compared the parallel DBSCAN performance between GPU and FPGA. The test data set is provided by the author of [6], which is 8-dimension data and single SPData is 256 bits. Frequency of FPGA is set to 100MHz in order to get 2 SPData from SDRAM in one cycle. Runtime is shown in Figure 8, the average speed up over GPU is 2.9x to 1.5x with 20 PEs. Meanwhile, the power consumption of GTX280 is 236w, and the Xilinx ML605 FPGA board is 15w. So our accelerator has a far better performance per Watt.

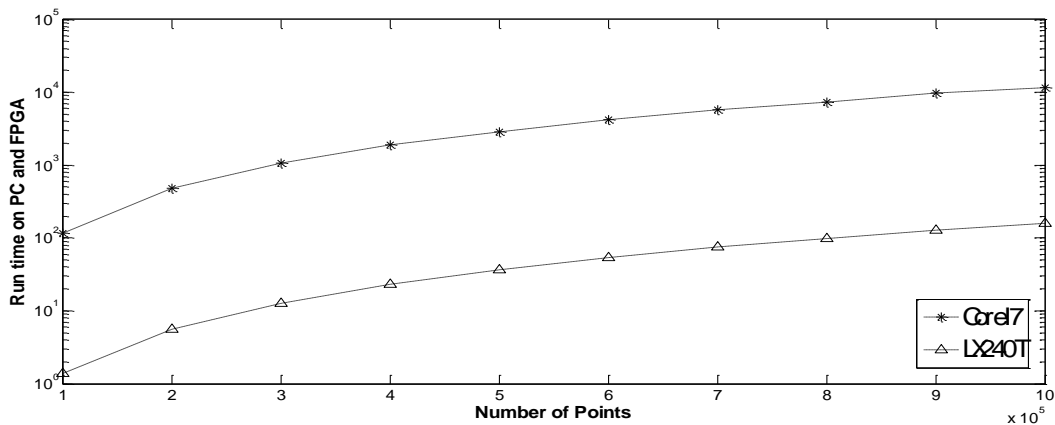


FIGURE 7 Runtime Comparison with CoreI7

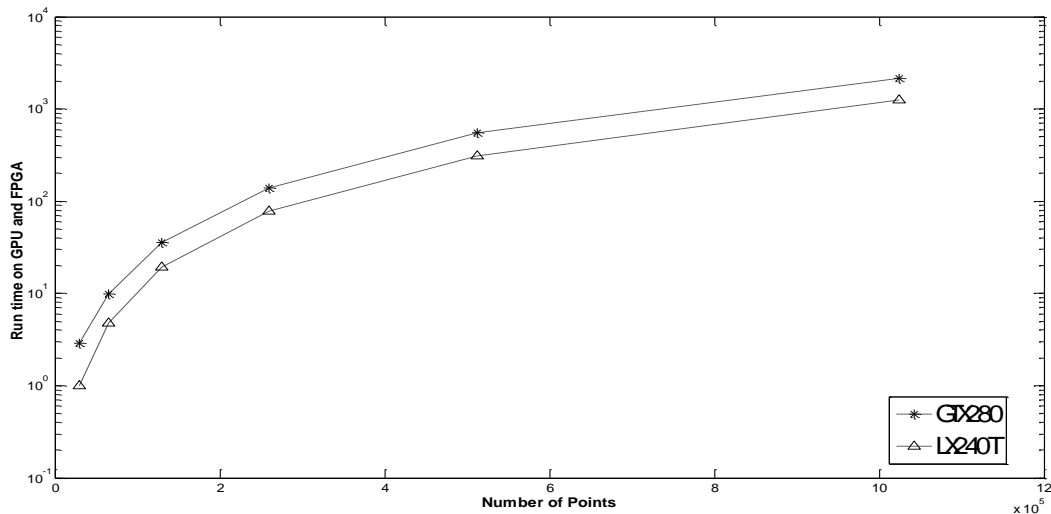


FIGURE 8 Runtime comparison with GTX280

6 Conclusions

In this paper, we have designed the hardware architecture for accelerating DBSCAN, which is a commonly used data mining algorithm. According to the deficiency of previous work, we propose a more robust collision check mechanism in parallel DBSCAN algorithm. Hardware

architecture we designed is based on the task-level and data-level parallelism. We have mapped this architecture to a FPGA-based accelerator. When compared with the PC and GPU implementation, experiment results show that our architecture can yield up to 86x and 2.9x speedup respectively.

References

- [1] Ester M, Kriegel H P, Sander J, Xu X 1996 A densitybased algorithm for discovering clusters in large spatial databases *Knowledge Discovery and Data Mining (KDD-96)* <http://dns2.icar.cnr.it/manco/Teaching/2005/datamining/articoli/KDD-96.final.frame.pdf>
- Xu X, Jager J, Kriegel H P 2002A fast parallel clustering algorithm for large spatial databases *High Performance Data Mining* 263-290 <http://dl.acm.org/citation.cfm?id=383194>
- [2] Arlia D, Coppola M 2001 Experiments in parallel clustering with dbscan *Proceedings of the 7th International Euro-Par Conference Manchester on Parallel Processing, ser. Euro-Par '01*. London UK: Springer-Verlag 326-31
- [3] Brecheisen S, Kriegel H-P, Pfeifle M 2006 Parallel Density-Based Clustering of Complex Objects *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science* 3918 179-188
- [4] He Y, Tan H, Luo W, Mao H, Ma D, Feng S, Fan J 2011 MR-DBSCAN: An Efficient Parallel Density-Based Clustering Algorithm Using MapReduce *The 17th IEEE International Conference on Parallel and Distributed Systems (ICPADS)* 473-80
- [5] Bohm C, Noll R, Plant C, Wackersreuther B 2009 Density-based Clustering using Graphics Processors *Proceedings of the 18th ACM conference on Information and knowledge management* 661-70
- [6] Cope Band Cheung P Y K, Luk W, Howes L 2010 *IEEE Transactions on Computers* 59(4) 433-48
- [7] Asano S, Maruyama T, Yamaguchi Y 2009 Performance comparison of FPGA, GPU and CPU in image processing *IEEE International Conference on Field Programmable Logic and Applications 2009 (FPL'2009)* 126-31
- [8] Estlick M, Leeser M, Szymanski J, Theiler J 2001 Algorithmic Transformations in the Implementation of K-means Clustering on Reconfigurable Hardware *Proceedings of the Ninth Annual IEEE Symposium on Field Programmable Custom Computing Machines 2001 (FCCM '01)*
- [9] Choudhary P, Choudhary A 2005 Design of a hardware accelerator for density based clustering applications. *Proceedings of the International Conference on Application-specific Systems, Architectures and Processors (ASAP), July 2005*
- [10] Narayanan R, Honbo D, Zambreno J, Memik G, and Choudhary A 2007 An FPGA Implementation of Decision Tree Classification *Proceedings of the IEEE International Conference on Design, Automation and Test in Europe (DATE), April 2007*
- [11] Baker Z K, Prasanna V K Efficient 2005 Hardware Data Mining with the Apriori Algorithm on FPGAs *Proceedings of the Thirteenth Annual IEEE Symposium on Field Programmable Custom Computing Machines 2005 (FCCM '05), 2005*
- [12] Baker Z, Prasanna V 2006 An architecture for efficient hardware data mining using reconfigurable computing system *The 14th Annual IEEE Symposium on Field Programmable Custom Computing Machines 2006 (FCCM '06) 2006*
- [13] Wen Y-H, Huang J-W, Chen M-S 2008 *IEEE Trans. Knowledge and Data Eng.* 20(6) 784-95
- [14] Sun S, Zambreno J 2011 *IEEE Transactions on Parallel and Distributed Systems* 22(9) 1497-505
- [15] Choudhary A N, Honbo D, Kumar P, Ozisikyilmaz B, Misra S, Memik G 2011 Accelerating data mining workloads: current approaches and future challenges in system architecture design. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1(1) 41-54
- [16] <http://www.rtreportal.org/> Oct 2013

Authors



Shaobo Shi , 08 1985, Beijing, China

Current position, grades: Ph.D

University studies: University of Science and Technology Beijing

Scientific interest: Micro architecture, VLSI design, Data mining, FPGA accelerator design

Publications: Shaobo Shi, Yue Qi, Qin Wang. Accelerating Intersection Computation in Frequent Itemset Mining with FPGA. The 15th IEEE International Conference on High Performance Computing and Communications, 2013.



Yue Qi, 11 1975, Beijing, China

Current position, grades: Ph.D, Lecturer

University studies: University of Science and Technology Beijing

Scientific interest: Micro architecture, VLSI design, Wireless Sensor Networks



Qin Wang, 01 1961, Beijing, China

Current position, grades: Ph.D, Professor

University studies: University of Science and Technology Beijing

Scientific interest: Micro architecture, VLSI design, Wireless Sensor Networks

Micro real-time pre-emption operating system for industry wireless sensor networks

Kanghong Duan, Hongxin Zhang, Shilin Song, Peigang Wang*

North China Sea Marine Technical Support Centre of State Oceanic Administration, 22 Fushun Road., Qingdao, China, 266033

Received 1 January 2014, www.tsi.lv

Abstract

Event-driven systems and thread-driven systems are two major design philosophy of operating system in wireless sensor networks. Systems based on multi-threaded are more timeliness than the event-driven systems, which can meet the requirements of time-critical tasks by means of task pre-emption, while systems based on event-driven are more energy efficient. Furthermore, μ COS-II is a classical system, which combines benefits in both systems. Therefore, our recent work we have shown that a micro real-time pre-emption operating system has been proposed on the basis of μ COS-II. First of all, a clear hardware abstraction layer (HAL) is given to combine the kernel and hardware in the system architecture. Moreover, this system is more capable of fitting both sensor network design goals of energy efficiency and timeliness. We are dedicated to modify the existing system from the scheduling strategy and data structure aspects, which lead to the performance of the modified system largely improved. Above all, the performance of our operating system is better than the original μ COS-II and TinyOS from task switch time, FLASH usage and RAM usage perspectives.

Keywords: Wireless Sensor Network, Operating System, Pre-emption, Improvement

1 Introduction

With the progress of sensor technology, communication technology and computer networking technology, WSN (Wireless Sensor Network) has been developed and improved rapidly. As one of the front fields, it brings us a lot of challenges, and the embedded operating system is one of them.

The embedded operating system that we research on must have the key functions as task scheduling, I/O management, timer management and so on. At the same time, it should pay more attention to the professional filed to meet the need of WSN.

Besides the general features in sensor networks, WSN has some its own features, which are important for the Industry WSN OS (Wireless Sensor Network Operating System).

The capability of the hardware is limited. The WSN node usually takes AA battery as power supply, and its microcontroller usually is 8-bit or 16-bit. At the same time, many applications and services may be limited by its memory spaces. These limitations require the operating system to be small and efficient [1].

The network should be large-scale, self-organized, dynamic and reliable. The amount of the node is large and the network topology changes quickly. Therefore, the operating system should have good methods to enhance its haleness, fault-tolerance and self-repair [2].

Wireless sensor network has very strong application relativity, and different hardware platforms and software systems are needed under different application

requirements [3]. So the operating system should be transplanted easily to meet the demands of different applications.

Many researchers has designed some WSN operating systems with high practical applicability. Based on different scheduling strategies, they can be divided into two kinds [3]:

- one is pre-emptive operating system oriented to hard real-time applications like MANTIS [4] and Contiki [5],
- the other is non-pre-emptive operating system oriented to soft real-time applications like TinyOS [6,7]. But these operating systems still have some disadvantages.

2 Study on μ COS-II

2.1 THE KERNEL OF μ COS-II

The kernel of μ COS-II is fully pre-emptive and real-time with multitask management; it can comparable on performance with most commercial kernels [8]. At the same time, it is only a Microkernel without many applications. Therefore, the architecture of it is easy to catch. FIGURE 1 shows the theoretical architecture of the kernel.

Most of the kernel's code is written by C language, only a few of the processor specific code is written by assembly language. And thanks to the clear hierarchical structure, the kernel gets the portal and scalable features which is crucial for WSN. And, the processor must

*Corresponding author - E-mail: dkhkidd@163.com

satisfy the following requirements [9]:

- 1) C compiler can generate reentrant code;
- 2) Interrupt can be disabled or enabled by C;
- 3) The processor must support interrupt and be able to generate timing interrupt;
- 4) The processor must support hardware stacks;

The processor must have instruction to load and store stack pointer and other registers to RAM.

2.2 SCHEDULE STRATEGY OF μ COS-II

μ COS-II uses the priority-based pre-emptive scheduling strategy. The base unit for scheduling in it is task. One task includes three main components: Program code in memory, TCB (Task Control Block) and stack space. Each task has its own priority and no task has the same priority while there are 64 priorities managed in μ COS-II. In order to ensure the schedule strategy, the kernel μ COS-II provides system services for tasks. FIGURE 2

shows the relationships between tasks and system services [9].

A task can request certain kinds of services of the kernel, then the kernel responses to corresponding requests. At the same time, the kernel executes the ready task with highest priority according to the state of the current task. The real-time is ensured by strategy of pre-emptive task scheduling.

2.3 THE MECHANISM OF TASK MANAGEMENT

After all codes downloading in the device, we can think that the whole system is ready to work. The first thing is starting the operating system and change the program to tasks.

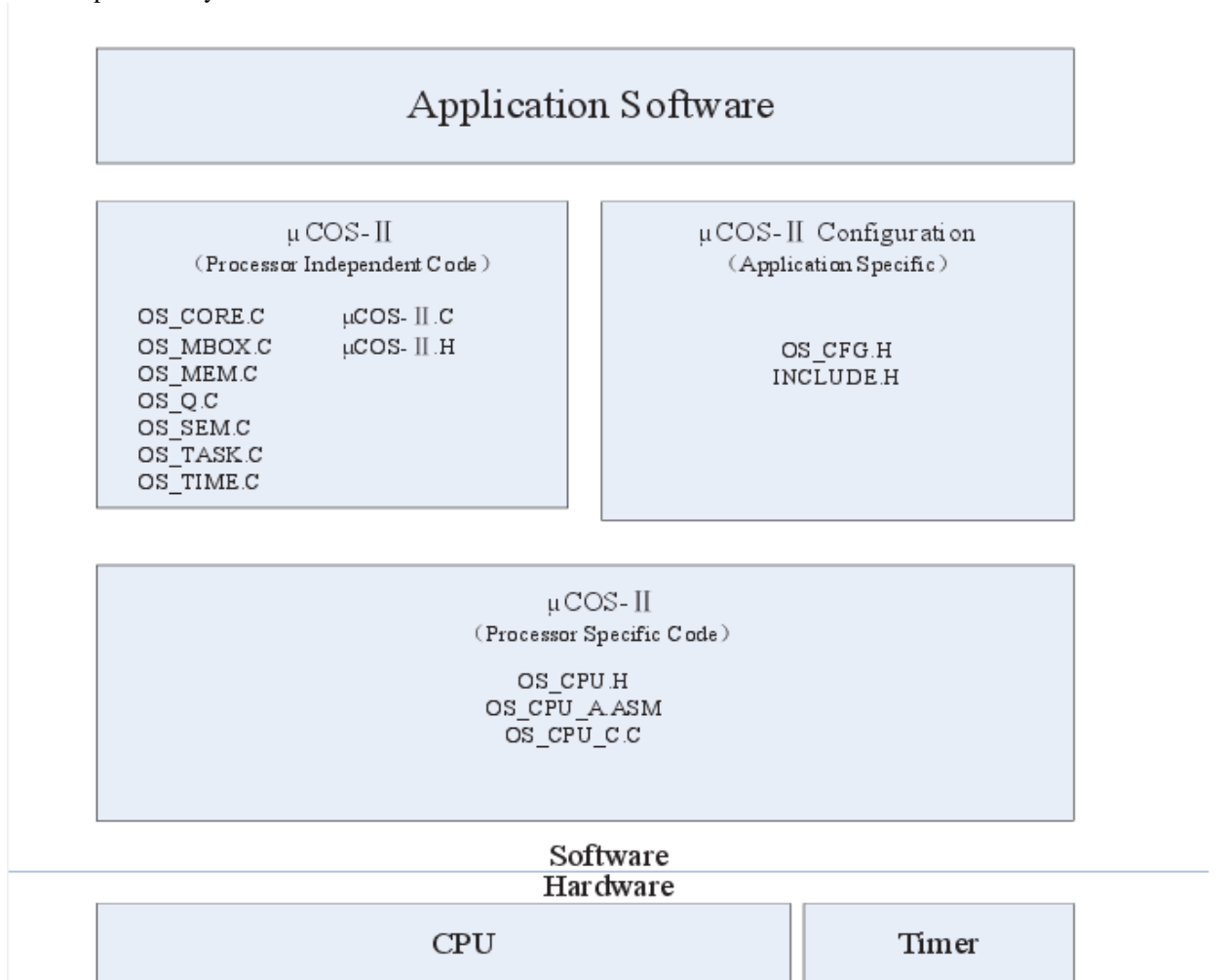


FIGURE 1 μ COS-II Hardware/Software theoretical Architecture

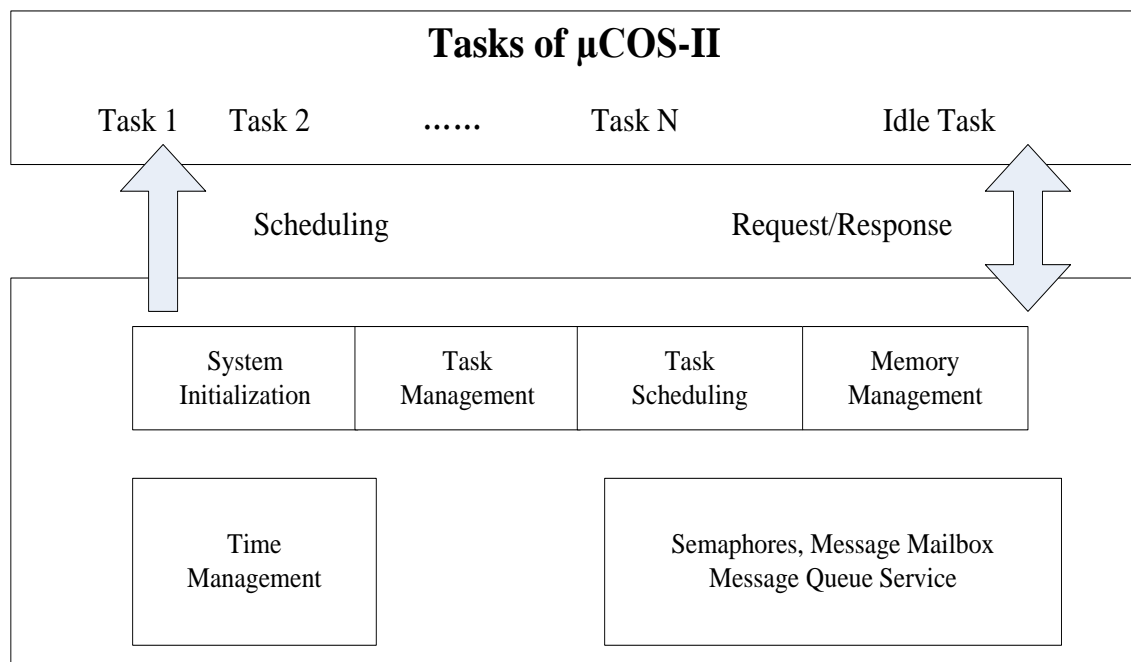


FIGURE 2 The relationship between tasks and system services

There are four states designed for every task: Dormant, Ready, Waiting and Running:

- 1) The dormant state is for tasks, which are deleted or not created. On this state, task does not have stack space and TCB while it cannot be scheduled. In other words, any dormant task is invisibility to the operating system.
- 2) The ready state is for tasks, which are ready for being scheduled. On this state, all the resources that task will use for running have been prepared well. And the task only need to wait until all the tasks who have the higher priority finish their work.
- 3) The waiting state is for tasks, which are waiting some resources necessary for running. On this state, task has stack space and TCB so that it can wait the necessary resources. After the resources are prepared, the waiting task will change its state to Ready.
- 4) The running state is for the task which is running at the moment. The running task has all the necessary resources and its priority is the highest.

In addition, there is ISR state for the interrupt. When a real-time call happens, the ISR will be triggered and OS will operate the new call first in order to ensure a strong real-time response.

The task management for μ COS-II is to ensure the following things:

- 1) Ensure the tasks in *Dormant* state occupying no RAM space.
- 2) Ensure the tasks in *Ready* state arranging in order.
- 3) Ensure the tasks in *Waiting* state preparing their resources in order.
- 4) Ensure the task in *Running* state running without disruption.

2.4 THE FEASIBILITY FOR TRANSPLANTING μ COS-II TO WSN

In the introduction we have given the three features which are important for WSN OS. Then we can research the feasibility based on the three features:

- 1) μ COS-II has a kernel perfect in function. But for WSN node, the memory capacity is a bottleneck for μ COS-II. Thanks to its scalable feature, we can improve some data structure and delete some useless functions for WSN in order to transplant μ COS-II.
- 2) μ COS-II has been used by hundreds of applications and its robust and reliable features are suitable for WSN.
- 3) Most of μ COS-II is written in highly portable ANSI C, with target microprocessor specific code written in assembly language. It determines the portable feature for μ COS-II. And the feature means we can transplant μ COS-II for WSN easily if we design a good platform for it.

The research tells us that μ COS-II is suit for WSN if the following things are done well:

- 1) Improve some data structure needing too much memory
- 2) Simplify some functions useless for WSN
- 3) Design a peripheral support platform for the μ COS-II

3 Target hardware platform for transplanting

Before improving μ COS-II for WSN, it is important to find a proper hardware platform for the research. In this paper, we use a normal WSN node hardware platform: ATmega128L micro-controller and CC2420 RF

transceiver chip. The main hardware resources are as follows:

CPU: 8bit micro-controller with 133 powerful instructions and on chips 2-cycle multiplier.

Memory: 4KB RAM and 128KB Flash ROM on chips.

RF chips: True single-chip 2.4GHz IEEE 802.15.4 compliant RF transceiver with baseband modem and MAC support.

4 Improving μ COS-II for WSN

The trend of WSN's design is for specifically applications rather than the general applications. And different

applications use different hardware platforms. Therefore, it is essential to change the details when you transplant μ C/OS-II to different platforms. Nevertheless, the principle for transplanting is universal.

4.1 IMPROVE THE ARCHITECTURE

For μ COS-II only provides a portable kernel, so the first thing for the improvement is to combine μ COS-II with the actual hardware. Then we can improve the original architecture of μ COS-II for the actual use. FIGURE 3 shows the improved architecture.

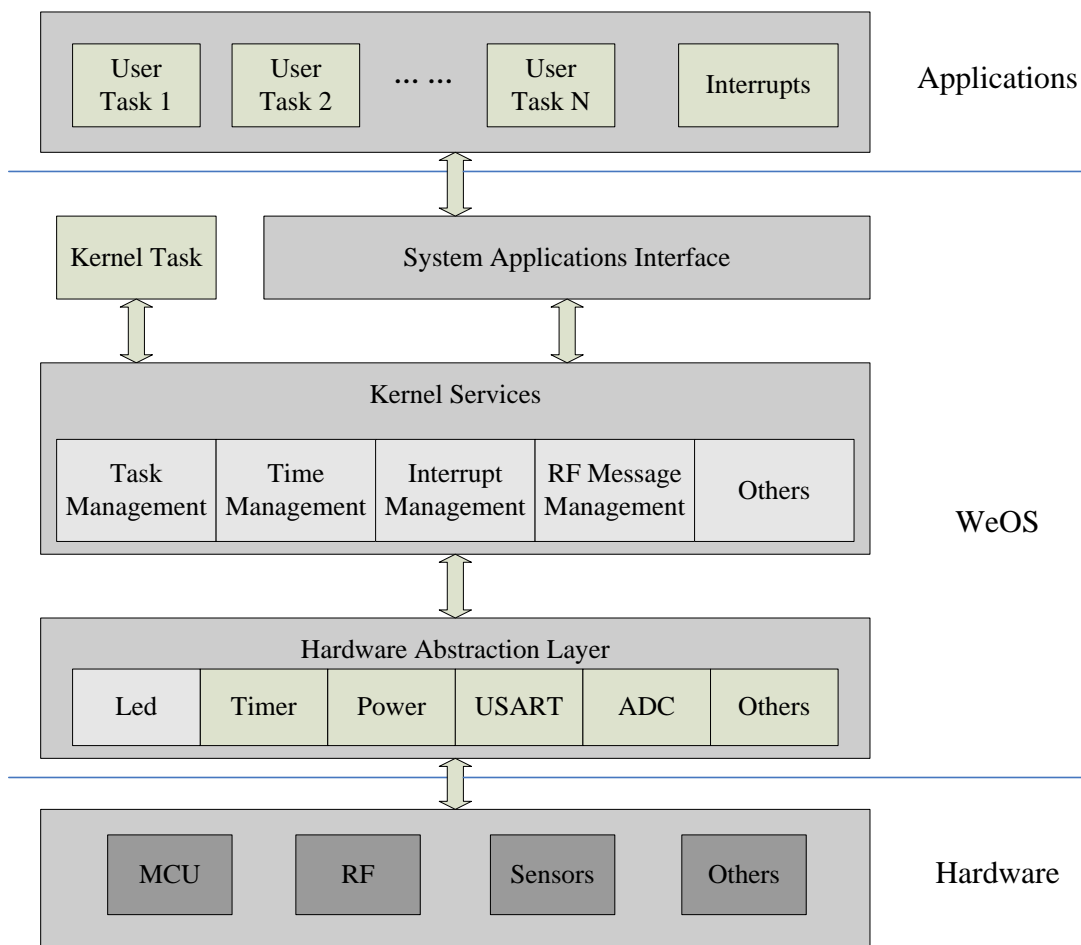


FIGURE 3 The improved architecture

There are two main improvements in the architecture:

- 1) Give a clear HAL (Hardware Abstraction Layer) to combine the kernel and hardware. And if any hardware should be added, moved or replaced, we only need to modify the corresponding code in the HAL without doing anything in the kernel and applications. By means of the improvement, we can transplant μ C/OS-II to the other platform more easier especially when we use a new kind of WSN node.
- 2) Add a new kernel task. As mention above, the capability of the hardware especially the power supply is limited. In order to reduce energy use, we

need to make the node sleep when there is no task need to be run. Therefore, we add an individual kernel task to finish the work. The kernel task has the lowest priority (always in Ready State) so that it is scheduled only when there is no user task ready. At the same time, there is no need to give any stack spaces to the task and the task can use the kernel service directly. By means of the improvement, the node can transform into sleep state by itself so that energy use will be reduced.

4.2 IMPROVE THE SCHEDULING STRATEGY

The scheduling strategy plays an important role in the performances of the real-time, power and reliability.

μ C/OS-II has a balanced scheduling strategy, but we can get a more suitable Scheduling Strategy for WSN by some improvements. FIGURE 4 shows the scheduling strategy.

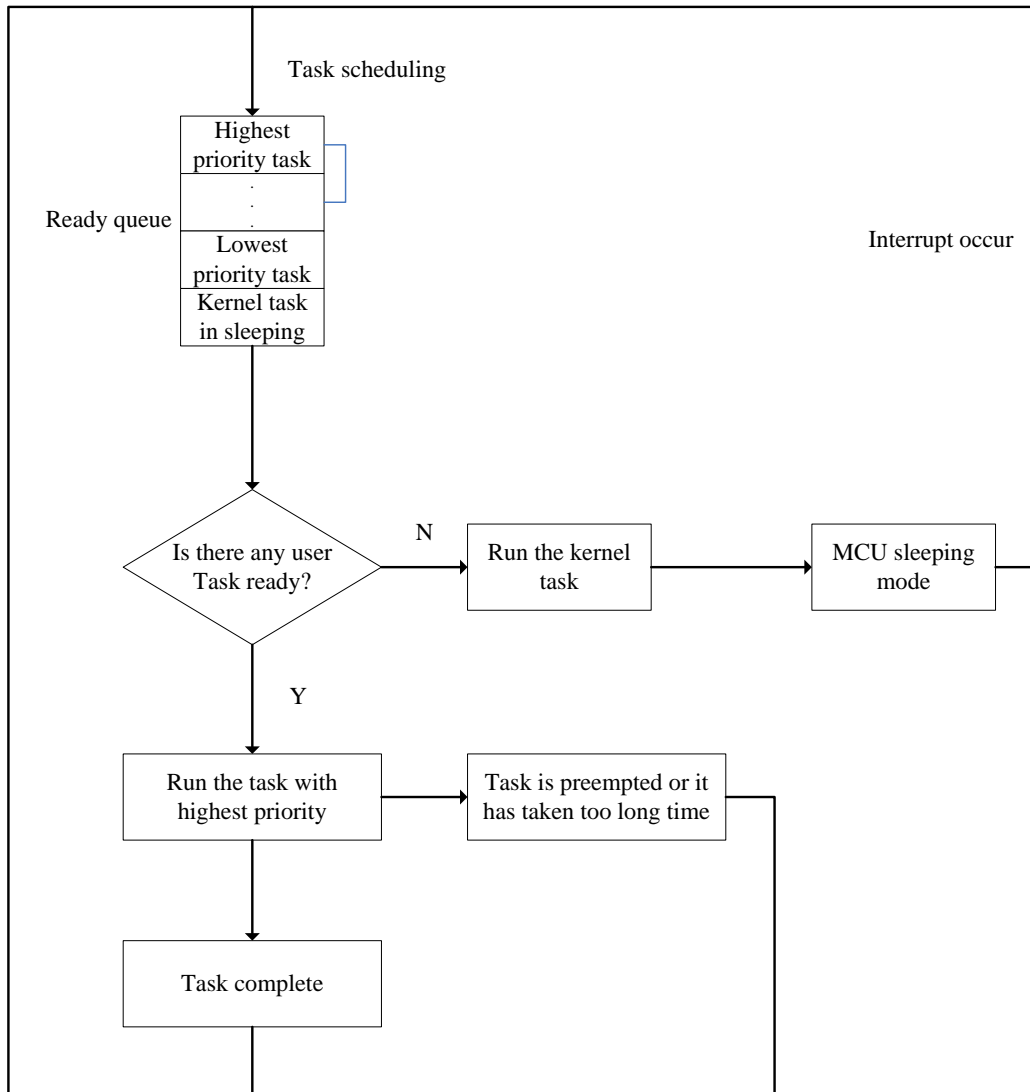


FIGURE 4 The improved scheduling strategy

There are two main important points in the improved strategy:

- 1) Modify the ready table. As mentioned above, we add a new kernel task with lowest priority and Ready State in order to make the node sleep when there is no user task ready. So we add a new judgement. When a procedure of task scheduling occurs, we first exam if there is any user task being in Ready State. If not, the kernel task will be scheduled and the MCU will get into Sleep Mode until an interrupt occurs.
- 2) Add a rotary-time scheduling method into the strategy. The original strategy takes the priority-based scheduling method and it has been proved excellent in

performance. But for actual applications especially in WSN, it is easy for users to create a high priority task with endless loop. It is clear that the problem will cause huge harm to the whole system even influence the other nodes. Therefore, we add a rotary-time scheduling method with hardware timers. And in the actual use, we first use the priority-based scheduling method to ensure the high priority task run smoothly, and the rotary-time scheduling method only ensure the task with endless loop can be stopped by the scheduler to make the whole system run well.

We use the following symbols to model the task of our operating system:

- S={s₁,s₂...s_n} is the set of task;
 - B={b₁,b₂...b_n} is starting time for each task;
 - L={l₁,l₂...l_n} is time slice for each task;
 - P={p₁,p₂...p_n} is priority for each task;
 - T={t₁,t₂...t_n} is the ending time for each task;
- Then we model the task of our operating system as (S, B, L, P, T).

For task s_i, its priority is p_i. And when s_i is called, the b_i will be recorded, and when the task is over (not include pre-empted by event), the t_i will be recorded. If the task

s_i, gets stuck in an infinite loop, when time slice l_i is over, s_i, will lose the control and change p_i to the lowest priority.

So we can get the worst waiting time for s_i:

$$T_{Waiting} = T_{Initial} + \sum_k^{i-1} l_k . \tag{1}$$

The improved OS simplifies many practical functions for applications in μC/OS-II and a new scheduling state method shown in FIGURE 5.

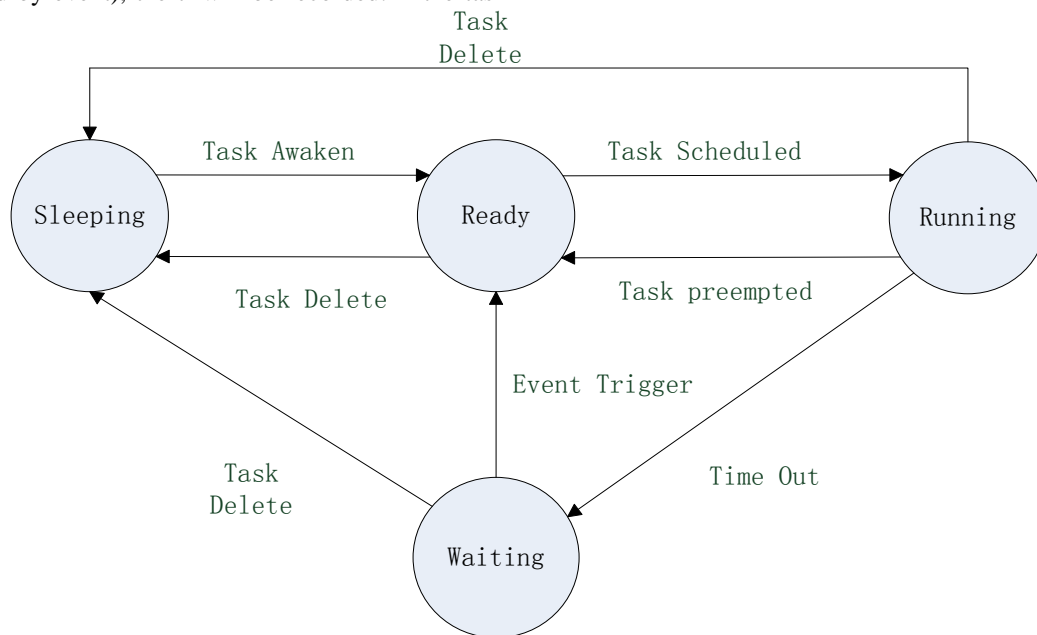


FIGURE 5 The improved task state

4.3 IMPROVE THE DATA STRUCTURE NEEDING TOO MUCH MEMORY

The limitation of memory capacity is the main problem for transplanting μCOS-II to WSN, especially the RAM space. So the main point that can be improved is to reduce the consume of RAM space for μC/OS-II.

The first thing we can do is to optimize the size and type of the global variable. In μC/OS-II, many global variables are defined as 32bits-integer data type which is usually useful for our WSN node. So we can change the 32bits-integer into 8bits-integer or 16bits-integer. At the same time, we can change the global variables which are not used frequently into extern global variables. So that these variables are not stored in RAM. And the RAM space can be saved a lot by the means above.

Another thing we can do is to enhance the table management. In WSN, the implementation of networking protocol usually uses many tables like routing table, neighbour table and so on. These tables usually are treated as temporary variables stored in RAM.

Therefore, we can enhance the management of these tables by limiting table size and setting public table RAM spaces to reduce the RAM consume.

The main members of RAM are global variables and temporary variables. So improving the size and type of these two variables is principle for saving more RAM space.

4.4 IMPROVE THE DATA STRUCTURE NEEDING TOO MUCH MEMORY

μCOS-II has many practical functions for applications, but in WSN, we don't need so many functions, so it's useful to delete some functions in order to get a more efficient kernel:

- 1) Simplify the procedure of creating a new create. We delete some operations, which are useful for system programming but useless for applications such as creating a pointer to the bottom of task stack. Then we can save some memory spaces and creating time by means above.

- 2) Simplify the mechanism of task communication. $\mu\text{C}/\text{OS-II}$ provides three kinds of methods for intertask communication: semaphores, message mailbox and message queue service. The functions of these three methods are similar. Therefore, we can choose one of them as our method. By this mean, we can get more space for applications instead of the operating system.
- 3) Change the method for handling interrupt. It is crucial for WSN node to response the external interrupt especially the interrupt from RF or sensors. Therefore, we get the interrupt handling out of the task state. Instead of that, we design the interrupt handling as a independent component in the whole system. By this way, we can get a well simplified kernel with a better capability for handling interrupt.

5 Performance analysis

In order to evaluate the improved $\mu\text{C}/\text{OS-II}$ especially its improvement for memory saving. We compiled TinyOS, original $\mu\text{C}/\text{OS-II}$ and our improved $\mu\text{C}/\text{OS-II}$ and downloaded them into our hardware platform. Then, we measured some key data to analysis the performance.

FIGURE 6 shows the task switch latency when a higher priority task is ready. From the result, we can see that TinyOS has the worst performance to ensure the higher priority task to be scheduled when it is ready. At the same, our improved $\mu\text{C}/\text{OS-II}$ has the similar performance with the original $\mu\text{C}/\text{OS-II}$. And they are about 5 times faster than TinyOS.

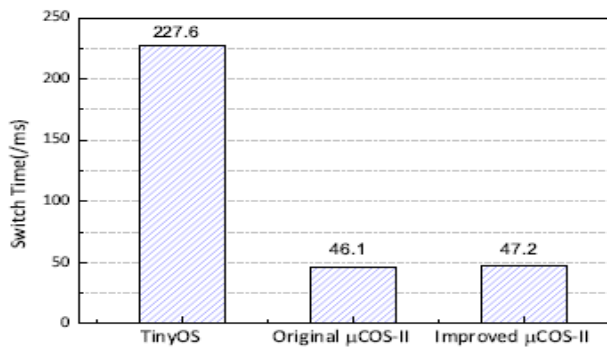


FIGURE 6 Task Switch Time

FIGURE 7 shows the Flash usage for each OS with different number and different size tasks. From the result, we can see that the original $\mu\text{C}/\text{OS-II}$ need about 40KB Flash space to hold the system itself. At the same time, our improved $\mu\text{C}/\text{OS-II}$ has similar performance with TinyOS, both of them only need less Flash space than half of the original $\mu\text{C}/\text{OS-II}$.

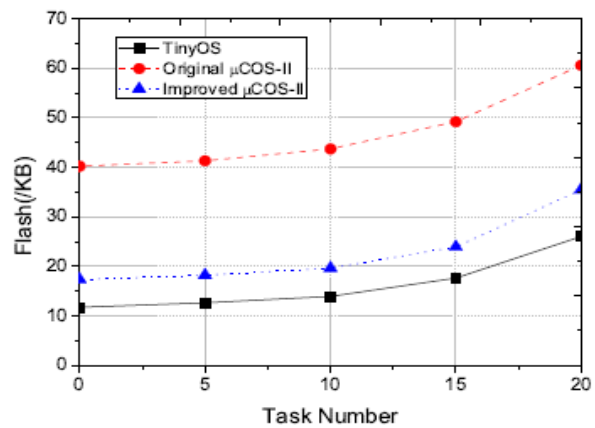


FIGURE 7 Flash Usage

FIGURE 8 shows the RAM usage for the original $\mu\text{C}/\text{OS-II}$ and our improved $\mu\text{C}/\text{OS-II}$ with different number tasks. From the result, we can see that the RAM space with different number tasks exhibit characteristic of linearity. The reason of the performance is that the RAM space is only relate to the initial system size and the size of task stack in $\mu\text{C}/\text{OS-II}$. The initial size for our improved $\mu\text{C}/\text{OS-II}$ is about half of the original $\mu\text{C}/\text{OS-II}$. At the same time, it is easier to manage the RAM space for each task in $\mu\text{C}/\text{OS-II}$ than in TinyOS.

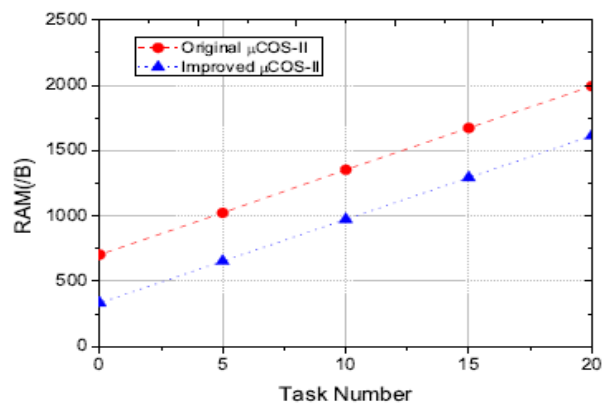


FIGURE 8 RAM Usage

Finally, we measure the performance for each OS with endless loop task. The result proves that our improved $\mu\text{C}/\text{OS-II}$ can stop the task when it takes too long time to occupy the MCU.

From the measure, we can conclude that our improved $\mu\text{C}/\text{OS-II}$ can provide a hard real-time scheduling for WSN with Relatively good space speeding.

6 Conclusion

This article analyses the limitations of WSN and studies the classic embedded operating system $\mu\text{C}/\text{OS-II}$ detailed. The paper brings forward a micro real-time pre-emption Operating System for industry Wireless Sensor Networks

and demonstrates the availability and performance of the improved operating system. It provides choices of more efficient system for the further study and application of Industry WSN.

References

[1] Akyildiz F, Su W, Sankarasubramaniam Y, Cayirci E 2002 *IEEE Communications Magazine* 40(8) 102-114
 [2] Minghai Y, Xiaoxiao Z 2007 *Wireless Communications, Networking and Mobile Computing* 1 2803-07
 [3] Wei D, Chun C, Xue L, Jiajun B 2010 *IEEE Communications Surveys & Tutorials* 12(4) 519-30
 [4] Abrach H, Bhatti S, Carlson J, Dai H, Rose J, Sheth A, Shucker B, Deng J, Han R 2003 MANTIS: System Support For Multimodal

Duan Kanghong, Zhang Hongxin, Song Shilin, Wang Peigang
 Networks of in-situ Sensors *The 2nd ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)* 1 50-9
 [5] Dunkels A, Gronvall B, Voigt T 2004 Contiki: A Lightweight and Flexible Operating System for Tiny Networked sensors *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks* 1 455-62
 [6] Levis A 2006 TinyOS: An Open Operating System for Wireless Sensor Networks *The 7th International Conference Mobile Data Management* 1 115-148
 [7] Ying L, Hongyi Z 2007 *Transducer and Microsystem Technologies* 26(03) 98-101
 [8] Jean J L 2002 *μCOS-II, The Real-Time Kernel* Press: Higher Education Press, Chapter 6 173-214 (in Chinese)
 [9] Kai L, Hongzhe X, Weiran X, Hongbo H 2005 *Journal of Communication and Computer* 38(6) 30-4

Authors	
	<p>Kanghong Duan, 04 1987, Zhucheng City, Shandong Province, P.R. China</p> <p>Current position, grades: engineer, North China Sea Marine Technical Support Centre of State Oceanic Administration University studies: Graduated from University of Science and Technology, master's degree in Computer Science Scientific interest: Internet of Things, Marine Information, Embedded operating system. Publications : More than 10 papers, 3 patents Experience: an engineer in North China Sea Marine Technical Support Centre of State Oceanic Administration</p>
	<p>Shilin Song, 11 1961, Qingdao City, Shandong Province, P.R. China</p> <p>Current position, grades: professor Senior Engineer, chief engineer, North China Sea Marine Technical Support Centre of State Oceanic Administration University studies: Graduated from Ocean University of China in 1983, Bachelor of Science in Marine Geology Scientific interest: Marine Geology, Marine Information, Wireless Sensor Network. Publications: more than 30 scientific research projects, more than 20 papers, 2 patents Experience: as an engineer, Senior Engineer and professor Senior Engineer in North China Sea Marine Technical Support Centre of State Oceanic Administration</p>
	<p>Hongxin Zhang, 01 1978, Qingdao City, Shandong Province, P.R.China</p> <p>Current position, grades: Senior Engineer, vice Chief of Centre, North China Sea Marine Technical Support Centre of State Oceanic Administration University studies: Graduated from Ocean University of China in 2000, Bachelor of Science in Physical Oceanography Scientific interest: Physical Oceanography, Marine Information, ship information service system. Publications: more than 10 scientific research projects, more than 20 papers, published, 6 patents Experience: an engineer and Senior engineer in North China Sea Marine Technical Support Centre of State Oceanic Administration</p>
	<p>Peigang Wang, 02 1962, Qingdao City, Shandong Province, P.R. China</p> <p>Current position, grades: professor Senior Engineer, Chief of Centre, North China Sea Marine Technical Support Centre of State Oceanic Administration University studies: Graduated from Ocean University of China in 1983, Bachelor of Science in Marine Biology Scientific interest: Marine Biology, Marine Information, Marine Geology Publications: more than 20 scientific research projects, more than 20 papers, 5 patents Experience: an engineer, Senior Engineer and professor Senior Engineer in North China Sea Marine Technical Support Centre of State Oceanic Administration</p>

Ground crack extraction in mining subsidence areas based on point cloud data

Ao Jianfeng^{1, 2*}

¹ School of Environment and Spatial Information, China University of Mining and Technology, Xuzhou 221116, China

² Jiangsu Provincial Key Lab of Resources and Environmental Information Engineering, Xuzhou 221116, China

Received 1 January 2014, www.tsi.lv

Abstract

During the underground mining of mineral resources, some ground points within a certain range above the working face will migrate correspondingly. The inconsistent horizontal migration of various ground points, due to different impacting time and degree, will lead to deformation of the ground surface, and when it reaches a certain degree, discontinuous damages to the continuous ground surface will occur, which are shown as cracks on the ground surface. In this paper, to extract planar cracks contained in the point cloud data, the point cloud data was firstly projected in the elevation direction for 2D processing and then a thickening algorithm for scattered point clouds was proposed. The point cloud data at different distances from the instrument was thickened block by block by choosing different thickening windows in order to highlight the crack data in the 2D image and to obtain the point cloud image applicable to the regular edge detection algorithm. Finally, the Canny operator was used to extract the edge information of cracks. In the end, the algorithm was tested in engineering projects and proved to be highly effective in extracting crack data.

Keywords: Crack Extraction, Point Cloud, Mining, Subsidence

1 Introduction

The underground mining activities in mine areas have tremendously impact upon the ground surface. Under serious circumstances, the impact upon rock stratum and surface will lead to unwatering of the underground water, landslide, debris flow and other geological disasters. In mining subsidence study, in addition to ground surface subsidence [1], cracks are also a common consequence induced by underground mining. Their development degree is an important criterion for identifying the damage degree of mine exploitation activities. In order to timely identify cracks and monitor their development, as well as to take an effective ground protection and underground mining control measures, the affected areas must be monitored in an effective manner.

In traditional mining subsidence monitoring, the crack data is collected by the total station based on several feature points at two terminals and the middle points of a crack to fit the whole shape of the crack. This method involves huge work and the practice of identifying feature points through human eyes easily gives rise to omission in massive data collection work and collected data is very limited. With the development of the surveying and mapping technology, the photogrammetry and remote sensing technologies have been successively applied to extract crack data. Atsushi et al. Ref [2] used a high-resolution camera to capture images and applied the integrated image processing technology to automatically extract micro cracks on the concrete block surface; Yang et al. Ref [3] designed the clustering target criterion

function based on inter-data manifold distance and optimized the function using the iterative method to monitor and extract data concerning road surface cracks; and Wei et al. Ref [4] combined the UAV image and TM image and used the ERDAS software to establish the knowledge model to extract crack data.

The introduction of 3D laser scanning technology provides a new means to acquire crack data. By using this method, the data may be collected quickly and sufficiently, thus enabling the crack data to be reflected in the point cloud in its completeness, and providing a reliable data bank for crack fitting. Moreover, the obtained true 3D coordinates are very helpful in crack positioning and measurement. Li et al. Ref [5] generated a profile based on the 3D laser scanning data to extract a crack on all profile lines. However, as the points on strip-shaped data clouds were selected as points to be interpolated, the elevation value at the interpolating points was smoothed to mitigate the abruptness on the profile lines, which would impact the effect of wavelet detection. Liu et al. Ref [6, 7] firstly projected the point cloud data onto the mean normal plane of the rock mass structural surface, and then, extracted the crack data on the structural surface by adopting the automatic fuzzy clustering algorithm [8]. As the ground cracks have various shapes, different approaches shall be adopted to guarantee a better extraction effect. In Ref [9], the author proposed a crack detection algorithm for stepped cracks in mine areas based on scanning line wavelet detection and carried out corresponding experiments, which produced good results. In consideration of different

*Corresponding author - E-mail: jfao008@163.com

development shapes, this paper mainly deals with the extraction method for another crack shape (i.e. planar cracks) in mine areas.

2 Theoretical foundation

2.1 GENERATION OF CRACKS

The ground surface subsidence induced by mining of underground coal resources is a spatial development process evolving with time [10, 11]. The impact of mining activities will lead to the migration of all ground points after being transmitted to the ground. The imbalance migration between all ground points will result in internal deformation of soil mass and after the tensile deformation goes beyond the deforming resistance of the ground, cracks occur [12]. The planar cracks generally occur outside the corresponding boundaries within the mining area and in front of the working face, as shown in Curve 3 in Figure 1 [13]. This area is the maximum tensile range of the underground subsidence basin. In the rear area of the working face advancement, as shown in Curve 4 in Figure 1, impacted by the mining activities, the surface soil within the region is compressed and deformed. With the progress of the mining program, the planar cracks start to close. However, as the earth surface cannot be restored to the original state and the cracks are closed to different extent, i.e. some planar cracks are not fully closed, with the final width smaller than that during the mining process, as shown in Figure 2(a); while some are over-closed, and form steps at the closing points, as shown in Figure 2(b).

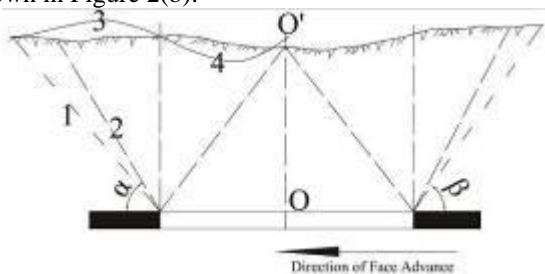


FIGURE 1 Model of surface migration. (1) Boundary of surface migration; (2) Crack surface; (3) Tensile deformation curve; (4) Compressive deformation curve; (a) Crack angle; (b) Boundary angle



FIGURE 2 Shape of cracks. (a) Shape of planar crack; (b) Shape of stepped crack

2.2 CANNY OPERATOR

In this paper, the edge detection based on Canny operator [14] was used to extract crack edges. The Canny operator used the first-order differential of the Gaussian function

and the non-maximum suppression and the “hysteresis” threshold value method to identify the maximum derivative for balancing the relationship between the anti-noise and detection and has showed good results.

The three optimization criteria for the first-order differential filter $h'(x)$ are as follows:

(1) Criteria of signal to noise ratio

$$SNR = \frac{\left| \int_{-w}^w G(-x)h(x)dx \right|}{\sigma \sqrt{\int_{-w}^w h^2(x)dx}}, \tag{1}$$

where, $G(x)$ is the Marginal function; $h(x)$ is the impulse response of the low-pass filter with bandwidth of W ; σ is the mean-squared error of Gaussian noise.

(2) Criteria of positioning precision

$$L = \frac{\left| \int_{-w}^w G'(-x)h'(x)dx \right|}{\sigma \sqrt{\int_{-w}^w h'^2(x)dx}}, \tag{2}$$

where, $G'(x)$ and $h'(x)$ are the first-order derivatives of $G(x)$ and $h(x)$ separately; L is the positioning precision of the margin.

(3) Criteria of single edge response

To ensuring the uniqueness of edge response, the mean distance of zero-cross point of the impulse response derivative of the detecting operator should be:

$$D_{zca}(f') = \pi \sqrt{\frac{\int_{-\infty}^{\infty} h''^2(x)dx}{\int_{-w}^w h''(x)dx}}, \tag{3}$$

where, $h''(x)$ is the second-order derivative of $h(x)$; f' is the image after detection.

Three criteria above depicted the criteria of edge detection. However, since the ability of anti-noise conflicts against the precision of edge detection, e.g. it will lower the precision of edge positioning after proper anti-noise procession; or to higher the precision of positioning, the ability of anti-noise will be influence, and some procession of compromising should be taken. The first-order of Gaussian function is the best compromising operator, which firstly take convolution on the original image for de-noising and then detect edge with derivation, the optimal detector Canny proposed is similar to the first-order derivative of Gaussian function.

The Gaussian is circular symmetry, so it is symmetrical along the marginal direction of Canny operator and antisymmetrical perpendicular to the marginal direction, so, aiming at the edge along the direction of most sudden change, the operator is very sensitive, but along the marginal direction is not sensitive.

Set 2-dimensional Gaussian function to be:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (4)$$

where σ is the distributional parameter of Gaussian function to control the smooth degree of image.

The optimal step edge detection operator based on the convolution $\nabla G * f(x,y)$, the intensity of the edge is $|\nabla G * f(x,y)|$, the direction is:

$$\rho = \frac{\nabla G * f(x, y)}{|\nabla G * f(x, y)|}. \quad (5)$$

Since the Gaussian function is unlimited, original templates are used to cut it off and get limited size N. Experiments show that when $N = b\sqrt{2}\sigma + 1$, the effect of detection is proper.

The specific realization is as follows:

First, since the Gaussian function is 2-dimension separable, two convolution templates of ∇G are divided into two 1-dimension filters:

$$\frac{\partial G(x, y)}{\partial x} = kx \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{y^2}{2\sigma^2}\right) = h_1(x)h_2(y), \quad (6)$$

$$\frac{\partial G(x, y)}{\partial y} = ky \exp\left(-\frac{y^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) = h_1(y)h_2(x), \quad (7)$$

where,

$$h_1(x) = \sqrt{k}x \exp\left(-\frac{x^2}{2\sigma^2}\right), h_1(y) = \sqrt{k}y \exp\left(-\frac{y^2}{2\sigma^2}\right),$$

$$h_2(x) = \sqrt{k} \exp\left(-\frac{x^2}{2\sigma^2}\right), h_2(y) = \sqrt{k} \exp\left(-\frac{y^2}{2\sigma^2}\right),$$

$h_1(x)=xh_2(x)$, $h_1(y)=yh_2(y)$, k is a constant.

Take convolution on the two templates with $f(x,y)$ separately:

$$E_x = \frac{\partial G(x, y)}{\partial x} * f; E_y = \frac{\partial G(x, y)}{\partial y} * f. \quad (8)$$

$$\text{Set } A(i, j) = \sqrt{E_x^2 + E_y^2}, a(i, j) = \arctan \frac{E_y(i, j)}{E_x(i, j)}.$$

Then $A(i,j)$ reflects the marginal intensity and $a(i,j)$ is the direction perpendicular to the edge.

According to *Canny's* definition, the central edge point is where the convolution of operator G_n and the image $f(x,y)$ hits its maximum in the direction of edge gradient. This law may be applied to determine whether a point is the edge point. A pixel is an edge point if meeting the following conditions:

- 1) Its boundary strength is greater than other pixels;
- 2) The directional difference between the point and its two neighbouring points in the gradient direction is smaller than 45° ;

- 3) The maximum boundary strength of the 3×3 neighbouring field centered on this point is smaller than a certain threshold value.

In addition, if the first two requirements are met simultaneously, the neighbouring pixels in the gradient direction shall be deleted. The point 3 is used to match the threshold value image composed by the edge points and the maximum gradient value within the range. The false edge points shall be deleted.

3 Flow of crack extraction

As there are huge amounts of noise data contained in the raw point cloud data, it is imperative to conduct corresponding de-noise processing before extracting ground cracks in order to eliminate non-ground points and obtain relatively smoothed point cloud data to facilitate the detection and extraction of crack data. The flow of crack detection is shown in Figure 3.

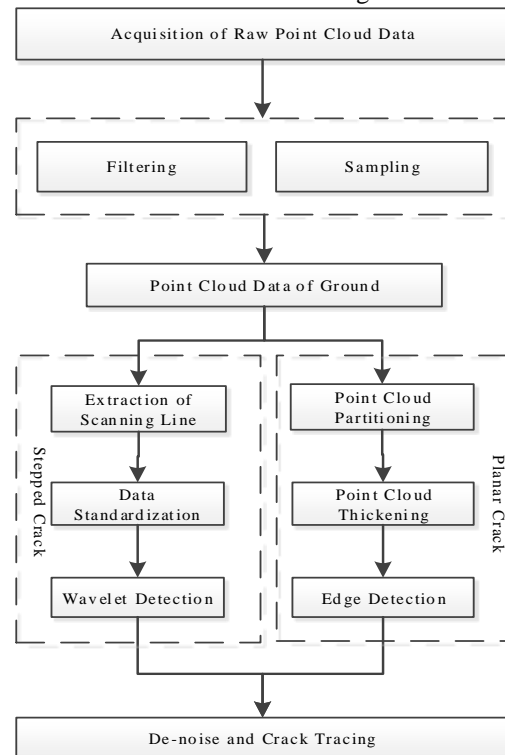


FIGURE 3 Flow of crack detection

For a planar crack, its data will be significantly missing in the point cloud data. The missing points are concentrated on one edge of the crack directly against the instrument. After the point cloud data is projected in the elevation direction, it may be observed clearly that there is a relatively large blank area in the crack area, as shown in Figure 4(b). If this point cloud data is processed graphically, i.e. the coordinates, after integer processing, serve as rows and columns of image pixels, and the elevation values of all points are converted into attribute values of pixels at the corresponding rows and columns (binarization processing is adopted in this paper, set values to 1 and blank to 0), the dimension reduction of point cloud data may be achieved. Then the edge

detection technique for 2D images may be used to extract cracks from the data.

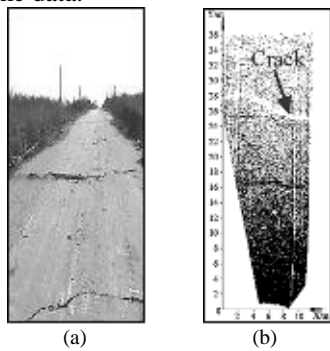


FIGURE 4. Photo of the ground and raw points of the ground. (a) Photo of the ground; (b) Raw points of the ground

However, when the 3D laser scanner is used to scan the surface of the mine areas, the space between the point cloud data will increase with the rising of distance, as shown in Figure 5(a). In addition, the instrument is short, and the angle of incident is small. These will gravely compromise the effective reflectivity of laser. The density of point clouds also reduces significantly with the increase of distance. Therefore, the point cloud data is sparsely distributed after the distance to the instrument reaches a certain limit, as shown in Figure 5(b). Therefore, the dispersion degree of non-zero pixels (i.e. pixels with attribute value $\neq 0$) in the point cloud image, which is gained through a graphical processing of point cloud data, is high and the space between non-zero pixels is large. The edge detection algorithm [17] based on image processing will fail if it is applied directly to point cloud data in this area. Therefore, it is necessary to thicken the 2D point cloud image first to narrow down the space between non-zero pixels in the non-crack area, and thus enabling the application of edge detection algorithm. Although this method will compromise the precision, with the increase of density of point clouds, the thickening windows may be narrowed appropriately to enhance the precision of crack extraction.

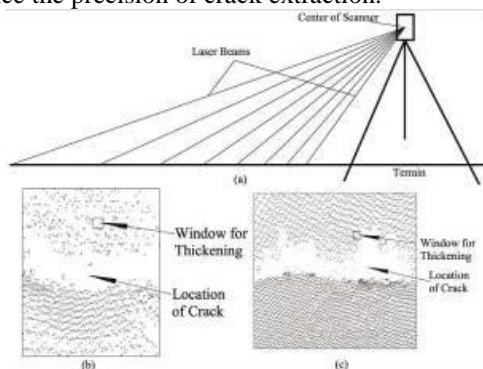


FIGURE 5. Thickening windows for point clouds with different density. (a) Style of scanning; (b) Points from long distance; (c) Points from short distance

In this paper, the moving window algorithm was adopted to thicken the point cloud image. The size of windows is dependent upon the scanning resolution at corresponding location of the point clouds, namely, the preset maximum point spacing at this location. The window size is normally 4 times the spacing, to ensure

there are certain points within the window when it is in the non-crack areas, as shown in Figure 5(b) and 5(c). With respect to every zero pixel in the point cloud image, first, move the window centre to the pixel, and then analyse the pixels in the window. As shown in Figure 6 that if there are scanning points within 4 quadrants of the window taking the zero pixel as centre, this zero pixel is within the non-crack area (see Figure 6 (a)); if at least one of 4 quadrants of the window has no scanning points, this zero pixel is within the crack area (see Figure 6 (b)). Therefore, during the analysis of the shifting window, if 4 quadrants of the window include non-zero pixels, set the pixels to 1, otherwise, keep the original value. After the pixels in the point cloud image have been thickened one by one, the edge detection algorithm based on Canny operator may be applied to the image. Through selecting different threshold values, the ground planar crack with different effects may be extracted. By combining the coordinates of all crack points and applying the algorithm in Literature [15, 16], the cracks may be traced.

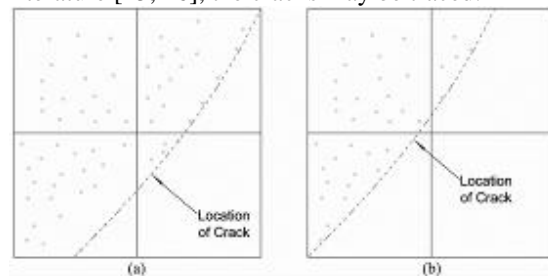


FIGURE 6 Point cloud thickening principle. (a) Points within non-crack area; (b) Points within crack area

4 Experimental verification

For a planar crack, first process the point cloud data graphically; then thicken the point cloud image obtained to highlight the crack area and smooth the non-crack areas; finally, detect and extract the crack edge with Canny operator. The detection results using the Canny operator with different threshold values are shown in Figure 7. It can be seen from the detection results that with the increase of the threshold value, the anti-noise ability of the crack detection effect based on the Canny operator is increasing till the noise data is completely deleted and the crack is extracted in its entirety.

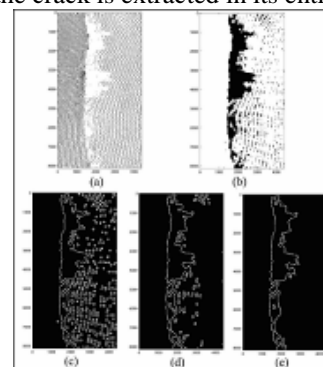


FIGURE 7 Detection of planar crack. (a) Raw points; (b) Image after thickening; (c) Threshold=0.15; (d) Threshold=0.55; (e) Threshold=0.95.

As the distribution of point cloud data obtained by the ground laser scanner is uneven and their density varies at different locations relative to the instrument, the thickening of ground point clouds shall be carried out block by block, and the thickening window shall be selected according to the local density of point clouds. The images of extracted cracks with different density are shown in Figure 8.

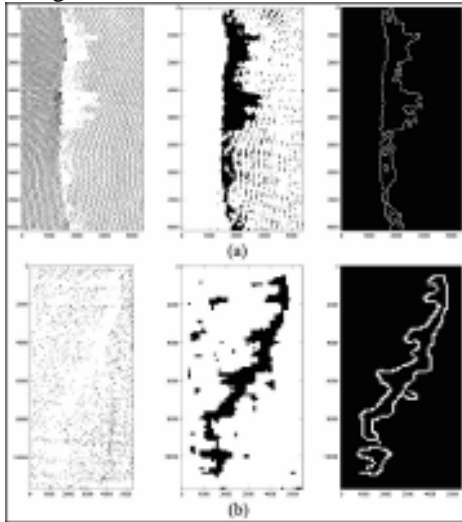


FIGURE 8 Extraction of cracks with different density. (a) Detection result of dense point clouds; (b) Detection results of sparse point clouds

5 Engineering practice

To verify the detection results obtained with the crack detection algorithm proposed in this paper, the ground point cloud, which was obtained by Trimble GX 3D from 5305(2) working face of Baodian Coal Mine Yanzhou mine area, was used to extract cracks, and the detection results were compared with the crack location surveyed with the total station. The undulation of ground surface is gentle within the area. See Table 1 for details of the 5305(2) working face.

TABLE 1 Basic information of 5302(2) working face

Working face length	Working face width	Average mining depth	Average ground elevation	Average coal seam thickness	Average mining rate	Coal seam dip angle
2400m	108m	305m	45m	8.7m	5.23m/d	8°

Figure 9 is the impression drawing based on crack detection results and an algorithm described in Literature [9].

It can be seen from Figure 9 that the crack data contained in the point cloud may be effectively extracted by combining two crack detection methods and the crack location thus extracted is consistent with that obtained with the total station. However, impacted by the point cloud density, the cracks with smaller width cannot be reflected in the point cloud, and therefore, cannot be extracted with the method described in this paper. To solve this problem, the point cloud density needs to be enhanced when data is collected to enable the smaller cracks to be better reflected in the point cloud data; moreover, some point cloud data is probably missing due

to oversized spacing between stations or shielding. The blank area is easy to be mistakenly extracted as a crack data when the detection algorithm is used to extract crack edges, and thus leading to problems shown in Figure 9. To solve this, the spacing between stations shall be reasonable and supplementary survey shall be conducted for shielded data.

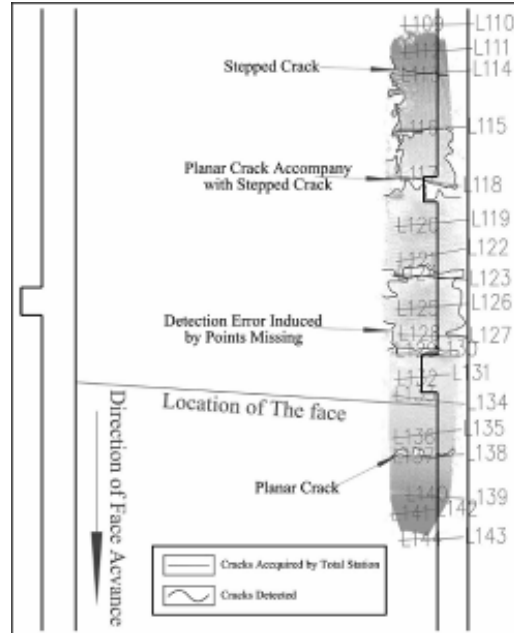


FIGURE 9 Impression drawing of crack detection

6 Conclusions

The 3D laser scanning technology may be applied in mining subsidence monitoring to gain abundant point cloud data, including subsidence information, crack information and features of characteristic constructions as well as other information conducive to the monitoring and study of mining subsidence. How to extract this kind of information is the key to the application of this technology. In this paper, the ground crack data in mine areas was successfully extracted, further pushing up the utilization efficiency of point cloud data.

First, the point cloud data was processed graphically to produce a 2D point cloud image; then the image was thickened and optimized to highlight the crack area, making the edge detection technology, which has been commonly applied in image processing, applicable to the detection and extraction of crack data.

Second, in this paper, the crack detection algorithm was also applied to field engineering projects to test its practicability. The good results further confirmed the reliability of this algorithm.

Acknowledgments

The research has been supported by a project funded by the Priority Academic Program Development of the Jiangsu Higher Education Institutions (PAPD, approval number: SA1202).

References

- [1] Dawei Z, Kan W, Ranli C, Liang L 2014 *Nat Hazards* **70**(2) 1197-208
- [2] Ito A, Aoki Y, Hashimoto S 2002 Accurate extraction and measurement of fine cracks from concrete block surface image *IECON 02*, 2202-07
- [3] Ruiui Y, Jianqiang N, Hongfei M 2011 *Computer Engineering* **37**(12) 212-4. (in Chinese)
- [4] Changjing W, Yunjia W, Jian W, Hui Z 2012 *Metal Mine* **436**(10) 90-2 (in Chinese)
- [5] Liang L, Kan W, Ranli C, Shu Z 2010 *Science of Surveying and Mapping* **35**(1) 165-6 (in Chinese)
- [6] Changjun L, Liuqian D, Dongya S 2011 *Chinese Journal of Rock Mechanics and Engineering* **30**(2) 358-64 (in Chinese)
- [7] Changjun L, Liuqian D, Baohui N, Shunfu Z 2011 *Chinese Journal of Rock Mechanics and Engineering* **30**(11) 2330-6 (in Chinese)
- [8] R E H, J H C 1998 *Int J Rock Mech Min* **35**(7) 889-905
- [9] Jianfeng A, Ming Z, Yanan X, Haixia Y 2013 *Metal Mine* **444**(6) 158-60 (in Chinese)
- [10] Guoqing H, Lun Y, Gengdi L, Fengcai J, Du H 1991 Mining subsidence science *China University of Mining and Technology Press (in Chinese)*
- [11] Youfeng Z, Kazhong D, Weiming M 2003 Mining subsidence engineering *China University of Mining and Technology Press (in Chinese)*
- [12] Jianrong K 2008 *Chinese Journal of Rock Mechanics and Engineering* **27**(1) 59-64 (in Chinese)
- [13] Xueyi Z 1996 *Journal of Xi'an University of Science & Technology* **16**(4) 295-9 (in Chinese)
- [14] Zhimeng X, Xiangdong G 2006 *Transactions of The China Welding Institution* **27**(1) 29-32 (in Chinese)
- [15] Yunpeng Z, Kan W, Yanan X 2012 *Coal Mining Technology* **17**(6) 10-3 (in Chinese)
- [16] Yunpeng Z, Kan W 2012 *Coal Science and Technology* **40**(8) 28-31 (in Chinese)
- [17] Ruijun G, Shenglei C, Jiakai W 2013 *Journal of Engineering Science and Technology Review* **6** (3) 10-5

Author

Jianfeng Ao , 06 1985, Xuzhou County, Jiangsu Province, P.R. China

Current position, grades: Ph.D studies in School of Environment and Spatial Information, China University of Mining and Technology, China.

University studies: BSc in Surveying and Mapping Engineering from China University of Mining and Technology in China. MSc from China University of Mining and Technology in China

Scientific interest: Mining Subsidence Monitoring, LiDAR Points Processing

Publications: more than 10 papers

Experience: participation in three scientific research projects

Expression model of multi-resolution 3D geographical space

Lou Ning¹, Zheng Xiaobo², Yang Yongchong^{1*}

¹ School of Geodesy and Geometrics of Xi'an University of Science and Technology, Xi'an, Shanxi, 7110054

² Construction, Office of Xi'an Aeronautical University, Xi'an, 100083

Received 6 February 2014, www.tsi.lv

Abstract

Aiming at improving deficiencies in multi-resolution two-dimensional expression of geographic information method, this paper studied method of multi-resolution 3D expression of geographical information, putting forward grid method to express different regional landforms separately and a method of expressing different regional ground objects on level of detail in three aspects as shape, texture and properties, plus an example to simulate and illustrate. Multi-resolution 3D expression of geo-information has broad application prospects in smart city and city planning, which has a certain exploring value for the massive information, for its expression and visualization.

Keywords: Multi-resolution, 3D visualization, Geographical Space Model, 3D terrain model, 3D ground object model

1 Introduction

The distribution of geographic entities in space is not uniform, and organization and expression of the geographic phenomena should be consistent with the geographical distribution. Therefore, the expression and visualization of geographic information will inevitably involve multi-resolution. In a digital environment, a geographic information can be expressed in different regions with different "scales" or resolutions. This kind of digital map is called multi-resolution digital maps. (Yang Yongchong, 2005, 2006)

In a two-dimensional digital map, where the geographic information is expressed in a multi-resolution, there are two aspects of disadvantages: the first is that the expression of the multi-resolution landforms is mainly described by a contour. The result of the expression is that this contour is intensive in the regions expressed in detail, while sparse contours correspond to weakly expressed regions. However, in the actual application of a contour line, intensive contour lines represent an area of a steep slope; sparse contour lines represent an area of moderate slope. If landforms are expressed by a contour of multi-resolution, doubts can be appeared (see, Figure 1).

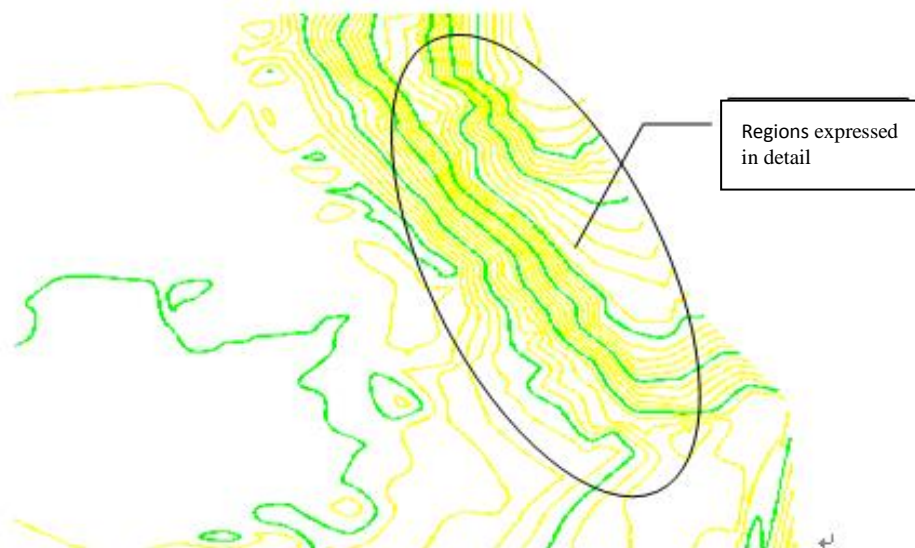


FIGURE 1 Landforms of different regions expressed by contour lines.

*Corresponding author - E-mail: 290272930@qq.com

In addition, in two-dimensional multi-resolution map, point, line, surface graphics are used to express the position information of a ground object. A height information and a spatial shape information of object cannot be accurately and intuitively expressed by a planar graph.

With the further expansion and improvement of the theory of spatial information, an expressing of geographic information by 3D spatial model becomes more intuitive, and more complete. In the 3D spatial model, referring to practical application, we can divide the entire expression area is into key areas and secondary areas. Then we make expression and visualization with different degrees in various areas, i.e. the multi-resolution 3D geographic spatial model.

This paper mainly studies the method of expressing terrain and ground objects in 3D multi-resolution.

2 Method of expressing multi-resolution regional terrain

There are three kinds of expressing geographic information in 3D multi-resolution: the 3D contour method, the triangulation method and the grid method.

2.1 THREE-DIMENSIONAL CONTOUR METHOD

The 3D contour (i.e. solid contour line) has obvious traces of steps. So it cannot express gully, steep ridge and other special landforms satisfactorily (see Figure 2). In addition, the feature of a closed contour is not suitable for the expression of multi-resolution terrain.

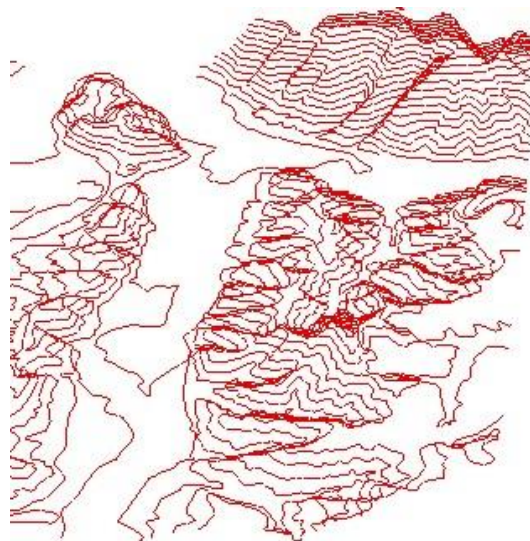


FIGURE 2 Three-dimensional contour method

2.2 THE TRIANGULATION METHOD

Triangulated irregular network (TIN) can reflect a complex terrain in a good way. It can also reduce the data redundancy. But because of the irregular shape and size of triangle, TIN has the same problem as contour

line since triangle is relatively smaller when expressing complex terrain, and relatively larger when expressing flat terrain triangular (see Figure 3) So it may cause misunderstanding just like contour line. Therefore, it is not suitable for the expression of multi-resolution terrain.

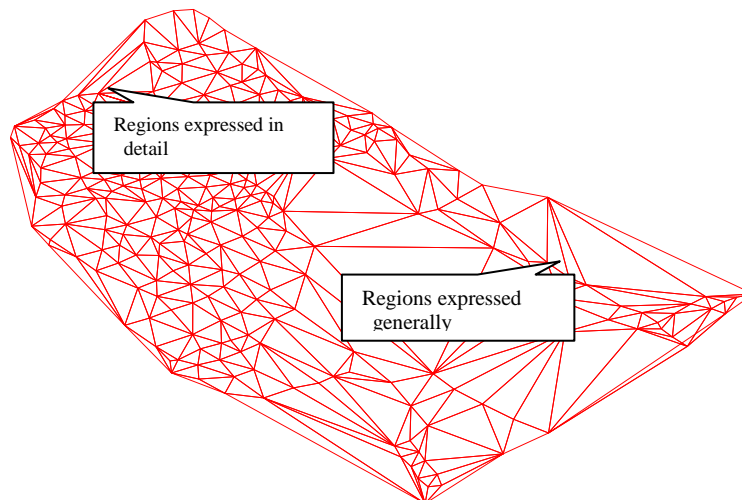


FIGURE 3 Geomorphology expressed by the triangulation method

2.3 THE GRID METHOD

The regular grid model is relatively regular. Big grids are used to roughly express landforms, and small grids are used to detailed express landforms. Regardless of differences in the complex of landforms, it is easy to

distinguish which areas are detailed expressed, and which areas are approximately expressed. So different sizes of grid can be used in multi-resolution to express different geomorphic information in different areas (see Figure 4).

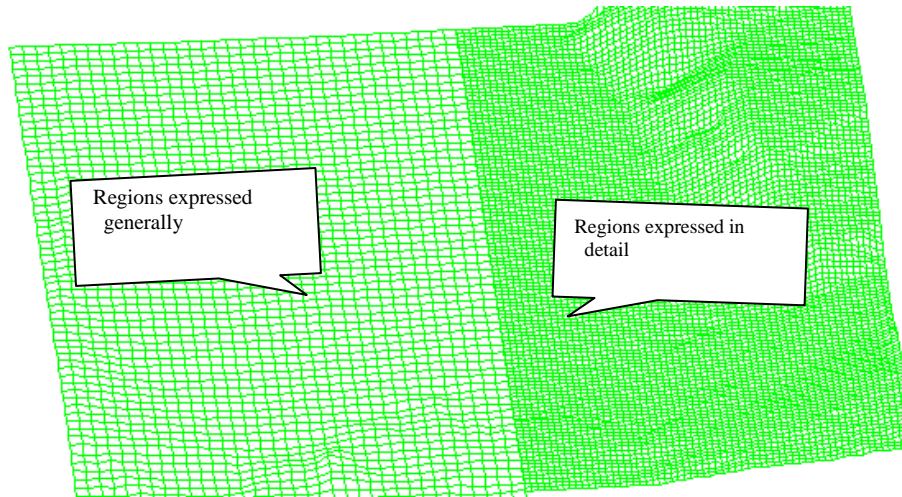


FIGURE 4 Using multi-resolution 3D regular grid method to express geomorphology

3 The method of expressing multi-resolution regional ground object

Spatial objects are complex and diverse. The expression method and the expression content are different in the key expression area and secondary expression area of multi-resolution 3D object. The differences lie in expression of its shapes, textures and properties.

3.1 MULTI-RESOLUTION OF SHAPE

Shape refers to the shape and contour of geographic entities. When doing visualization of geographical spatial information, shape is an important visual variables. map

interpreter can recognize different qualitative and quantitative specifics of point, line and surface elements according to the shape.

In geographic information expressed in 3-dimensional multi-resolution, object shape represents for its true shape in key areas. Therefore, the expression is very detailed and accurate, obtaining highly precise information. In addition, in the secondary areas, the approximate expression of object shape is often done through combining small-area ground object for large-area ground objects, and method of abandoning the small-area objects, simplifying object contour and other combing method (see Figure 5)



FIGURE 5 Shape information of ground objects expressed in 3-dimensional multi-resolution

3.2 MULTI-RESOLUTION OF TEXTURE

Different entities in a geographical space have different details on its surface. Through these details, which are also textures we can recognize different geographic entities, judging differences between entities of the same shape. And by "sticking" texture, the geospatial entities generated can be realistic and three-dimensional.

In the expression of Multi-resolution 3D object, real texture of geospatial entities are used to express the detail information of ground objects in key areas. This expression is consistent with the objective entity, laying stress on the authenticity of the object. In the secondary areas, general textures in the texture database are used to generally express information of geographical entities, reflecting general properties of the object, stressing its existence.

2.4 THE MULTI-RESOLUTION OF PROPERTY

The different Geospatial entity has different properties information is an important feature to cognize ground objects and to distinguish between them.

In the expression of 3D multi-resolution geographic information, properties values of spatial entities are expressed in different degree. With the regard to ground objects in key areas, geographic entity properties information is very detailed and accurate, enabling map interpreter to obtain the most detailed entity information. In the secondary areas, the entity's overall information and properties information are only briefly described. In key areas such as the area on the left side of Figure 5, you can not only query to properties information of village, but query to properties information of every household in the village. But in the secondary area on the right side of Figure 5, only the properties information of village can be queried.

4 Conclusions



Geographical terrain and ground objects are of complexity and diversity. In key areas and secondary

areas where geographic information are expressed in multi-resolution 3D, small grids are used to express important regional geomorphology in detail and big ones secondary areas in general, regardless of complexity degree of the terrain and ground objects. In the key areas, shape and texture of 3D models look close to geography spatial entities and properties are also detailed. In secondary areas, houses are expressed in an abstract way. After the cartographic generalization, and by using general building textures and expressing only the main information concerning the properties, map interpreters are provided with a compendium of information from its shape, showing the macro feature of the generally expressed regions. This is the geographical space model expressed in multi-resolution 3D.

The method of multi-resolution 3D expressing of geographic information has a wide application prospect in engineering and city planning. As long as there are key areas and the secondary area s in graphics areas, you can use the method of 3D multi-resolution to express geographic information.

References

- [1] Yongchong, Yang, Zhida Guo 2005 *Bulletin of Surveying and Mapping* 7 13-16
- [2] Yan Zhou, Qing Zhu, Duo Huang 2006 *J. Science of Surveying and Mapping* 5 15-18
- [3] Lan Guo 2002 *J. Bulletin of Surveying and Mapping* 5 10-11
- [4] Chengming Li, Jizhou Wang, Zhaoting Ma 2008 *The Principle and Method of 3D Geo-spatial Framework of Digital City* 1 Beijing: Science Press
- [5] Lan Guo, Yongchong Yang, Hongtao Tang 2009 *J. Geotechnical Investigation & Surveying* 37(4) 67-71
- [6] Yongchong Yang, Nan Hu, Lan Guo 2011 *J. Bulletin of Surveying and Mapping* 1 75-77
- [7] Junqiao Zhao 2013 *J. Acta Geotactica et Cartographica Sinica* 1
- [8] Qingyuan Li, Zongjian Lin, Chengming Li 2002 *J. Science of Surveying and Mapping* 6 17-21
- [9] Xin Yang, Guoan Tang, Xuejun Liu, Fayuan Li, Shijie Zhu 2009 *J. Acta Geographica Sinica* 9
- [10] De Florian L, Magillo P, Puppo E 2000 *Compressing Triangulated Irregular Networks. GeoInformatica* 6 Berlin-Heidelber: Springer

Authors	
	<p>Ning Lou, 10 1978, Xian Yang, Shan Xi province, China</p> <p>Current position, grades: MSc, lecturer Scientific interest: Digital map; GIS and its application</p>
	<p>Xiaobo Zheng, 03 1986, Xian Yang, Shan Xi province, China</p> <p>Current position, grades: assistant Scientific interest: digital map; the visualization of geographical information</p>
	<p>Yongcheng Yang, 10 1966, Gaolan County, Gansu province, China</p> <p>Current position, grades: Professor Scientific interest: Geographic Information Engineering and cartography</p>

Cloud database dynamic route scheduling based on polymorphic ant colony optimization algorithm

Chen Qing^{*}, Yong Zhong, Liuming Xiang

Chengdu Inst. of Computer Applications, Chinese Academy of Science, People's South Road, Section 4 Chengdu, China

Received 1 March 2014, www.tsi.lv

Abstract

Big Data era spawned the development of Cloud database. As a database, which need easily scale out, how to quickly find the available nodes are focuses of the study. Ant colony algorithm is based on bionic optimization algorithm and has the characters of smart searching, global optimization, robustness, distributed computing and easily combined with other algorithms, but the algorithm is prone to premature convergence, making the results often caught local optimum. According to this, polymorphic ant colony algorithm was proposed which combined with a Cloud database; the algorithm can quickly and reasonably find the nodes in Cloud environment, reducing the load of routing, thus greatly improved the Cloud database's ability of scaling out.

Keywords: Big Data, Cloud database, Ant colony algorithm, premature convergence, scaling out

1 Introduction

As the representatives of the Sensor systems just as RFID are widely used, global data are growing explosively. These data have characters of complexity, spontaneity and randomness. How to effectively manage these big data has become a big challenge. Cloud database as a data storage system, has been studying for long. Cloud database are different from traditional distributed database and static database, how to find the most suitable nodes as soon as possible in order to improve database's ability of scale-out is the focus of research.

2 Cloud database

Cloud database system consists of a number of sites gathered together. These sites also called nodes, which are joined together in the communication network, each node is an independent database system, and they have their own database, the central processor, terminals, and their respective local database management system. Therefore Cloud database system can be seen as a series of joint centralized database system [1], and this system also known as architecture of share-nothing (SN). They logically belong to one system, but on the physical structure are distributed. Big Data Cloud database as an effective management system that has the following characteristics:

- 1) Highly Scalable: Cloud database nodes provide dynamic expansion and contraction capabilities, users can adjust number of nodes in the entire system according to needs, and is convenient for user's resource allocation and management.
- 2) High Reliability: each data node at least have multiple copies, a copy of the data nodes automatically

translate to the master node and provide data access services when a master data node crashed.

- 3) High Stability: Cloud database system can keep running as long as there has only one node.
- 4) High Efficiency: data dynamically balanced allocation algorithm, fully coordinated, balanced each storage server's storage pressure to ensure that the data on the storage server stress basically average; various databases' parallel scheduling algorithm, fully coordinated each node within Cloud systems to work together, to provide users with microsecond response support.
- 5) Low Cost: Cloud database can be used as an extremely low-cost computer system to build Cloud node, system construction costs can be controlled at a lower level. Some explanations above Cloud database are distributed, dynamic, global balanced and scalable database.

At present, the design of Cloud Databases scaling out have two main ideas. One is the database based on key/value structural, which belongs to non-relational databases, and the database system can easily sale out due to no correlation between the data, this design greatly enhanced the level of scalability, but because of lacking of transaction management, it can only be used in a particular scene, such as Social Networking Services(SNS). The other idea of design is based on traditional distributed relational database (DDBS), because this system retains the management of transaction, how to improve the ability of scale-out to adapt Big Data environment is the difficulty in designing.

In this paper, we did not focus on the transaction level, but focus on the node-finding algorithm when scaling out which all the design of CloudDB would face to in Cloud environment.

^{*}Corresponding author - E-mail: qing33dd@aliyun.com

3 Basic ant colony algorithm models

Ant colony algorithm is typically used for solving complex combinatorial optimization problems. In solving the problems of different properties, the ant colony algorithm model definitions are different. We take TSP [2], which has m nodes as an example to illustrate the basic ant colony algorithm model. m nodes TSP problem is to find through m nodes each time and finally back to the starting point of the shortest path.

Let n be the number of ants in ant colony, d_{ij} (i, j = 1, 2, ..., m) is the distance between node i and node j, $\tau_{ij}(0)$ is the concentration of pheromone of node i and node j at time t. At initial time, concentration of pheromone is the same on each path, set $\tau_{ij}(0) = C$ (C is constant). Ant k (k=1, 2, ..., n) In the process of movement, concentration of the pheromone in each path decides the direction, $p_{ij}^k(t)$ is the probability of Ant k transfer from the node i to node j at time t, which is calculated as (1):

$$p_{ij}^k = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)}{\sum_{s \in tabu_k} \tau_{ij}^\alpha(t)\eta_{is}^\beta(t)} & j \notin tabu_k \\ 0, & j \in tabu_k \end{cases} \quad (1)$$

$tabu_k$ (k=1, 2, ..., n) is the set of nodes which ant k has passed. In the beginning, $tabu_k$ has only one element, which is beginning node, along with the evolution, elements in $tabu_k$ are increasing. As time goes on, the pheromone in the path gradually disappeared. Parameter (1-Q) represents the degree of pheromone's volatilization, all the ants to complete a circulation; each concentration of pheromone is adjusted according to the formula (2),

$$\left. \begin{aligned} \tau_{ij}(t+1) &= \rho * \tau_{ij}(t) + (1-\rho) * \Delta \tau_{ij} & \rho \in (0,1) \\ \Delta \tau_{ij} &= \sum_{k=1}^n \Delta \tau_{ij}^k \end{aligned} \right\} \quad (2)$$

$\Delta \tau_{ij}^k$ is the concentration of pheromone in path ij left by ant k in circulation, and $\Delta \tau_{ij}^k$ is the sum of concentration of pheromone in path ij left by ant k in circulation.

Dorigo has given three different models, referred to ant cycle system, ant quantity system, ant density system; the difference lies in their different calculation expression of $\Delta \tau_{ij}^k$ are different.

In the model of ant cycle system,

$$\Delta \tau_{ij}^k = \begin{cases} Q/L_k, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases} \quad (3)$$

L_k is the distance walked by ant k in this circulation.

In ant quantity system and ant density system, $\Delta \tau_{ij}^k$ means:

$$\Delta \tau_{ij}^k = \begin{cases} Q/d_{ij}, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases} \quad (4)$$

$$\Delta \tau_{ij}^k = \begin{cases} Q, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases} \quad (5)$$

In the three models above, the first model uses global information, the other two use local information, so the model of ant quantity system is usually used as basic model. Parameters can use experimental methods to determine the optimal combination and also can be obtained by evolutionary learning. The calculation ceased when the evolutionary trend is not obvious.

4 Ant colony algorithms

P The literature [3] shows that real ant colony society is organized, and the work is divided. The ants in ant colony are divided into labour ants, scout ants, soldier and worker ants, etc. They carry out their duties, and also collaborate, forming as an organic one, and each ant contact with each other with hormones secreted, the class of hormones are divided into warning class, attracting class, convening class, marking region class etc. While a single ant is small, but by forming a group, it reflects the high degree of organization and society, so that the entire ant colony has extraordinary abilities. Based on this, Xu Jingmin [4] et proposed the based on polymorphic ant colony, where the "polymorphic" refers to the ant society has a variety of the state and pheromone.

4.1 COMPARING WITH OTHER SEARCHING ALGORITHMS

Comparing with evolutionary computation [5-7], ant colony algorithm is similar within two aspects; First, two algorithms used group to express solution of the problem; second, the new groups are created by the knowledge related to old groups. And main difference between two algorithms are that in evolutionary computation knowledge of all issues are included in the current groups, while in ant colony algorithm, all the knowledge in trail of pheromones.

And comparing with the simulated annealing algorithm (SA) [8], the search strategy is almost the same in essential. In SA, the process of calculating the energy E_i at state i in a solid is same with ants in a "travel", they are both sampling the solution of space; "annealing" and "secrete pheromones" are both using accumulated information to enhance the subspace search; and "Metropolis Standard" and "random state transition rules" are both trying to use the algorithm to escape from local optima, and accept the worse within a certain range, and search for a new sub-space.

So From the perspective of ant colony algorithm, we can see that: the scale of ants actually affect the frequency of updates, so if the scale is large, the gap between different state become widen when updating the pheromone; and when the scale is small, only a small amount of pheromone will be updated in order to avoid premature .and moreover, because both algorithms are essentially the same, some improvements and variations in SA can be used directly on the ant he ant colony algorithm to improve its performance.

Ant colony also can be considered as connection system, and the most representative example of connect system is neural networks (Neural Network, referred NN) [9, 10]. From a structural point of view, ant colony algorithm and neural networks usually have a similar parallel mechanism; every state ant visited corresponds to the neural networks of neurons. Ant itself through the neural network can be seen as concurrent input signal. Ant colony algorithm learning rule can be interpreted as an acquired rule, that the signal are more strength in good quality solutions than in bad solutions.

4.2 SHORTAGE OF EXISTING ANT COLONY ALGORITHM

Although the ant colony algorithm has many advantages, but there are also some drawbacks of this algorithm. Compared with other methods, the algorithm generally require a longer search time, and complexity of the algorithm can reflect that , and also the algorithm is prone to stagnation , just like that when searching at a certain time, all the solutions become consistent, and cannot search further in solution space for a better solution.

In the ant colony algorithm, the ants are always dependent on other ants feedback to reinforce learning , rather than to consider their own experience, that herd behaviour, easily lead to precocious, stagnation , so that the algorithm converges slowly . Based on this, scholars have proposed a new ant colony algorithm to improve the algorithm. Dorigo M proposed an ant colony algorithm called Ant-Q System; German scholar Stutzle T and Hoos H proposed an algorithm called "max-min ant system" (MAX-MIN ant system, MMAS). Wu Qinghong [5] inspired by mutation operator of the genetic algorithm, and used a reversal mutation mechanism, and proposed an ant colony algorithm having mutation characteristics, and which is the earliest improvements in ant colony algorithm made by domestic scholars. Since then there have been many scholars making improvement in ant colony algorithm , such as an ant colony algorithm with characteristics of sensation and perception [11] , self-adaptive ant colony algorithm [12], an ant colony algorithm based on pheromone diffusion [13], an ant colony algorithm based on hybrid behaviour [14], and an ant colony algorithm based on pattern-learning [15, 16] .

4.3 POLYMORPHIC ANT COLONY ALGORITHM

In polymorphic ant colony algorithm, the ant colony society is divided into three categories: scouting ants,

working ants and searching ants. Scouting ant's task is to use the node as starting point for local reconnaissance, and mark the results by hormones in order to leave the message for coming ants; searching ant's task is doing global searching, choosing the next node by hormones and searching until finding the best (shortest) route, and mark the route for working ant; working ant task is to feed back the food to the nest on the marked route which is shortest one. In algorithm designing, because working ant not responsible for routing, so we only need to control scouting ant and searching ants' pheromone.

Algorithm of scouting ants: put m ants in m nodes, and using node as centre to scout other (m-1) nodes, and combined results with MAXPC used as pheromone, marked $s[i][j]$, which means the route from node I to node j. Calculated as follows:

$$s[i][j] = \begin{cases} d/d_{ij}, & \text{if node j in the scope of nodei's MAXPC} \\ 0, & \text{else} \end{cases} \quad (6)$$

d_{ij} is the shortest distance from node i to other m-1 nodes. And according to this, the amount of information on each path the initial time set as follows:

$$\tau_{ij}(0) = \begin{cases} c*s[i][j], & \text{if } s[i][j] \neq 0 \\ c*d/d_{ij}, & \text{else} \end{cases} \quad (7)$$

d_{ij} is the longest distance from node i to other m-1 nodes, and C is concentrations of pheromone on the roads at initial time, and pheromone is support for calculating of transition probability and provide support for adjusting the concentration of pheromone on each path.

Algorithm of searching ants: probability of ant k (k= 1, 2 ... n) moving from node i to node j at t time, $p_{ij}^k(t)$ is:

$$P_{ij}^k = \begin{cases} \frac{\tau_{ij}^\alpha(t) \eta_{ij}^\beta(t)}{\sum_{s \neq tabu_k} \tau_{ij}^\alpha(t) \eta_{is}^\beta(t)}, & j \neq tabu_k, \text{ and } s[i][j] \neq 0 \\ 0, & \text{else} \end{cases} \quad (8)$$

According to this formula, the searching ant reduces the scale of the search due to the pheromone.

If all the ants complete a circulation, each concentration of pheromone will be adjusted as follows:

$$\tau_{ij}(t+1) = \begin{cases} \rho*\tau_{ij}(t)+(1-\rho)*\Delta\tau_{ij}, & \text{if } s[i][j] \neq 0 \\ \rho*\tau_{ij}(t), & \text{else} \end{cases} \quad (9)$$

$\Delta\tau_{ij}$ is the sum of the pheromone left by ant moving on the path ij in a circulation. $\Delta\tau_{ij}^k$ is the pheromone left

by ant k moving on the path ij from time t to time $t+1$ in a circulation, and $\Delta\tau = \sum_{k=1}^K \tau_{ij}^k$ is as follows:

$$\Delta\tau_{ij}^k = \begin{cases} Q*(d/d_{ij})/L_k, & \text{if ant } k \text{ passed } ij, \text{ and } s[i][j] \neq 0 \\ 0, & \text{else} \end{cases} \quad (10)$$

The processes of Polymorphic ant colony algorithm can be described as follows:

- 1) Initialize Q , C and maximum evolution algebra;
- 2) Put m scouting ants in m nodes, each ant scouts other $m-1$ nodes and calculates the pheromone using formula(6), and put the results in $S[i][j]$; Two different styles of referencing are in common use: the Harvard alphabetical system and the Vancouver numerical system. Use the Vancouver numerical system as described below.
- 3) Using formula (7) to initial information in each path;
- 4) Set $NC = 0$;
- 5) Randomly selecting each scouting ant's initial position and put the position in table tabu;
- 6) Using formula (8) to calculate each scouting ant's position to be transferred;
- 7) Calculating the value of the objective function ants search L_k ($k = 1, 2, \dots, n$), and record the best solution at current;
- 8) If the algebras or the solution have not got any answer, go to step 10); otherwise modify the concentration of pheromone according to formula (9);
- 9) Set $\tau = 0$, and set table tabu Null, $NC \leftarrow NC+1$, go to step 5) ;
- 10) Output the optimal answer.

5 Experiment

5.1 EXPERIMENTAL MODEL

We choose NS-2 which developed by South California University as platform [17], and use the throughput and delay of transmission time as indicators, and select the algorithm OSPF(Open Shortest Path First) [18], SPF (Shortest Path First) and BF for comparison.

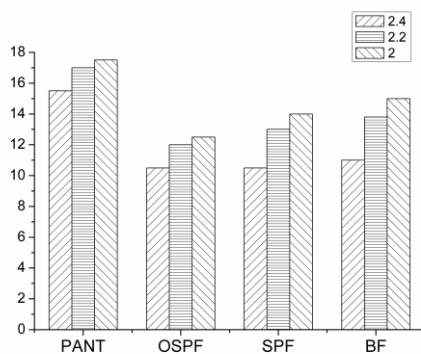


FIGURE 1 Average throughput

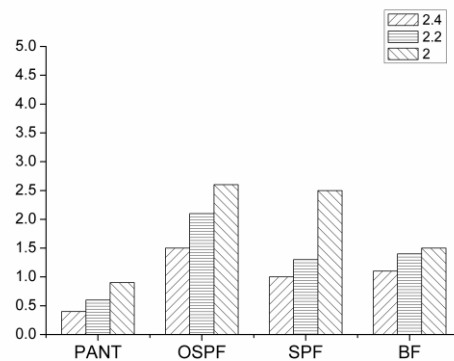


FIGURE 2 The Performance of packet delay

From FIGURE 1 and FIGURE 2, we can find that the polymorphic ant colony algorithm performances better than the other three in throughput and delay of transmission time. And through the experiment, we can also find that in the algorithm based on polymorphic ant colony, if the inspired factor α is too small, convergence will be slow and, easily fall into local optimum; if inspired factor α is too large, pheromone's weight will be heavy in scouting, and causes premature convergence. If expected factor β is too small, the ant colony will lead into purely random searching, and difficult to find the optimal solution; if β is too large, the speed of convergence be faster, but convergence tends to be bad. If pheromone evaporation factor ρ is too large, previously searched path had possibility of researching, and it will affect the algorithm's randomness and global searching capability; if pheromone enhancement factor Q is larger, accumulation of pheromone will be faster, and it can enhance capability of the positive feedback in searching, and fasten the convergence; when Q is too large, the algorithm of the global search capability will deteriorate, easy to fall into local optimal solution, and cause the loops.

6 Conclusions

In this paper, we combined polymorphic ant colony algorithm with Cloud database. The algorithm improved previous basic ant colony algorithm, and meanwhile solving the problem of node dynamic plugging when the Cloud database needs to scale out. In experiment, the throughput and delay of transmission time were used as indicators to evaluate algorithm's capability; and using NS-2 which developed by South California University as platform. The results were compared with other similar algorithms, and shows Cloud Database can quickly find the node in Big Data's environment using polymorphic ant colony algorithm.

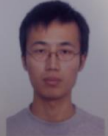

Acknowledgments

This research is supported by Science and Technology Department of Sichuan Province.

References

[1] <http://soft.ccw.com.cn/it/> / Oct 2013
 [2] Du D P 1990 *Logistics Graph Theory* Buaa. Press: Beijing 240-54
 [3] Wu J, Wang CH L 1995 *China Ant* Chinese Forestry Press: Beijing 1-12
 [4] Xu J M et al 2005 *Journal of University of Science and Technology of China* 35(1)
 [5] Wang Z Z, Bo T 2000 *Computational Evolution* National University of Defence Technology Press: Changsha
 [6] Li Q M, Kong J S, Lin D, Li SH Q 1995 *The Basic Theory and Applications of Genetic Algorithms* Science Press: Beijing

[7] Liu Y, Kang L S, Chen M P 1995 *Parallel Numerical Algorithms: Genetic Algorithms* Science Press: Beijing
 [8] Kang L S, Xie Y, Liu SH Y, Luo ZH H 1995 *Non-numerical Parallel Algorithms: Simulated Annealing Algorithm* Science Press: Beijing
 [9] Zhang N Y, Yan P F 2000 *Neural Networks and Fuzzy Control* Tsinghua University Press: Beijing
 [10] Jin F 2000 *Fundamentals of Neural Computational Intelligence Methods* Southwest Jiaotong University Press: Chengdu
 [11] Chen L, Qin L, Chen H J, Xu X H 2004 *Journal of System Simulation* 33(2) 983-9
 [12] Wang Y, Xie J Y 2002 *Journal of System Simulation* 14(1)
 [13] Huang G R, Cao X B, Wang X F 2004 *ACTA Electronica SINICA* 32(5)
 [14] Hu X B, Huang X Y 2005 *Control and Decision* 20(1) 69-71
 [15] Xiao Y S, Li B Y, Wu Q D 2004 *Control and Decision* 19(8) 885-88
 [16] Mi J S, Leung Y, Wu W Z 2005 *Int J Gen Syst* 34(1) 77-90
 [17] www.isi.edu/nsnam/ns/ Oct 2013
 [18] Moy J 1991 OSPF version 2. RFC 1247

Authors	
	<p>Chen Qing, ChengDu, China</p> <p>Current position, grades: ChengDu, PhD studies University studies: Chinese academy of Science Scientific interest: Big data, Cloud Database Publications: A Seamless Software Development Approach using DCI</p>
	<p>Liuming Xiang, born in November, 1989, ChangSha, China</p> <p>Current position, grades: ChengDu, Master studies University studies: Chinese academy of Science Scientific interest: Big data, Cloud Database</p>

Matters of satellite queuing network design in K_a-band for Republic of Kazakhstan

A Aitmagambetov^{1*}, Yu Butuzov², S Torekhan¹

¹ International IT University, Department of Computer Technologies and Telecommunications, Manasa 34, 050040, Almaty, Kazakhstan

² Almaty University of Power Engineering and Communications, Baitursynov 126, Almaty, Kazakhstan

Received 1 February 2014, www.tsi.lv

Abstract

This work was carried out within the framework of research opportunities of using K_a - band satellite communication systems in the Republic of Kazakhstan. The paper deals with the multi-beam coverage in Kazakhstan (the distribution of beams in area and determine their capacity), as well as evaluation of the main parameters of subscriber channels. The need for this research was due to the fact, that the design of multibeam network for Kazakhstan is important to consider a distinct uneven distribution of the population (or potential users), low average density (about 6 persons / sq km.) and a fairly significant differences rain intensity in some areas of the territory.

Keywords: satellite networks, K_a-Band, multi-beam antennas, bandwidth, signal-code constructions

1 Introduction

Recent years, the accelerated development of Ka-band as a global trend is observed in satellite communication. A considerable number of works devoted to the review of existing and planned satellite queuing systems [1-3], operating experience [4] and the results of research on the optimization of their parameters [5-7]. Before application of multi-beam technology the provision of broadband services to the mass consumer in Ka - band was considered less profitable because of the need for a super-cheap VSAT - terminals. With an advent of queuing systems, which are based on the technical application of multi-beam receiver and transmitter onboard antennas, the above mentioned problem has been solved.

The present work was carried out within the framework of research opportunities for application of Ka-band satellite communication systems in the Republic of Kazakhstan. The work is devoted to the study of multi-beam coverage in Kazakhstan (the distribution of the beams of the grounds and the determination of their capacity), as well as the evaluation of the main parameters of subscriber channels. The need for these studies is due to the fact that the design of multi-path network for Kazakhstan it is important to take into account the pronounced uneven distribution of the population (potential customers), lower average density (about 6 persons / sq km.) [8] and is quite a significant difference on the territory of the individual zones of intensity rain.

The most densely populated region, where 1 sq.km for about 20 people, is the South-Kazakhstan region, and the

most sparsely populated region with a minimum density of 2.3 people per 1 sq. km. km - is Aktobe region.

Levels of rainfall intensity exceeding the 0.01% of the year duration change from 10 mm / h in the western and central regions of up to 30 mm / hour in the East Kazakhstan region [9].

When choosing a geostationary satellite orbital position 58.5 E stage by stage was taken into account the fact that Kazakhstan submitted to the ITU in the Ka-band for this position 3 applications [9] (the latter KAZSAT-1R with a priority date of 11.14.2012 and was valid until 30.03.2018).

2 Formation of the working area of the satellite

Among three options forming of working area and satellite distribution capacity in the beams considered in [5, p.1] for Kazakhstan (considering the uneven population density) is set equal distribution of the beams of the grounds and the uneven distribution of capacity in the beams.

In accordance with the documents of ITU and CEPT for projected satellite network, the frequency bands listed in Table 1 were selected.

TABLE 1 Frequency bands for Kazakh satellite

The transfer hub (CES)	The transfer ST
29.0 – 29.7 GHz	30.0 – 30.5 GHz
The reception hub (CES)	The reception ST
18.3 – 19.0 GHz	19.2 – 19.7 GHz

Note: The hub – central earth station (gateway); ST – subscribe (user) terminal.

* Corresponding author - Tel: +7(701)-204-09-21
E-mail: altayzf@mail.ru

In Kazakhstan satellite repeater is supposed to apply separate receiving and transmitting multi-beam antennas (MBA), which will optimize the antenna maximum gain (MG), lower level of the side lobes (LSL) and reduce the cross-polarization radiation, and more precisely, sustain

mutual consistency of the viewing zones at the reception and transmission. The distribution of the beams on the territory of the Republic of Kazakhstan is shown in Figure 1.

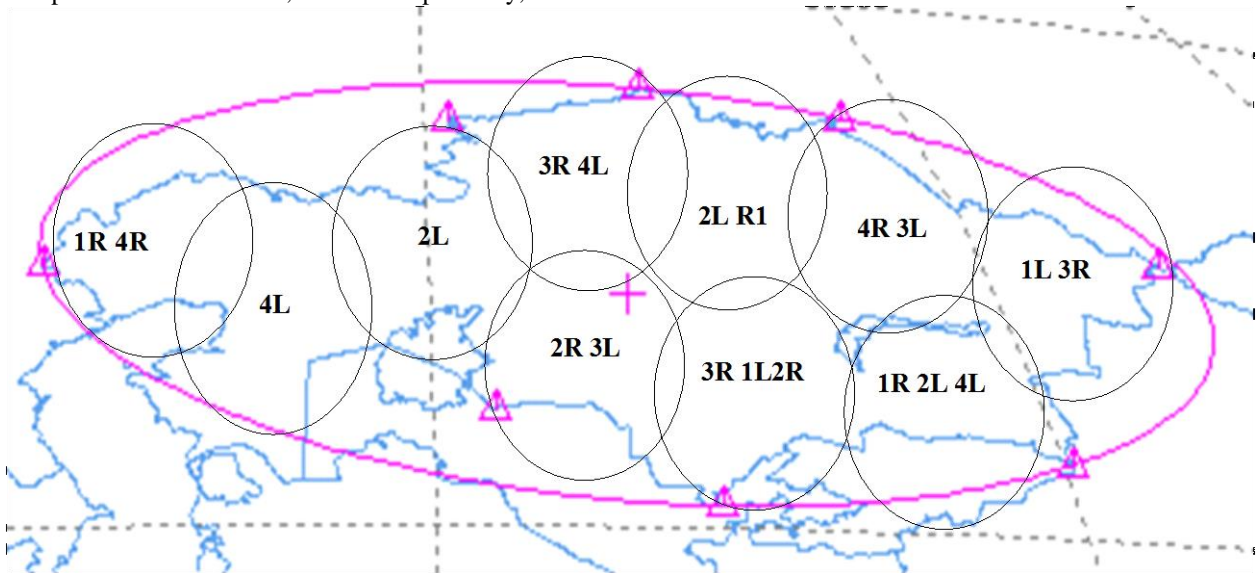


FIGURE 1 The uniform distribution of beams on the territory of the Republic of Kazakhstan and the unequal distribution of capacity in the beams

Working frequency range of 500 MHz, of the band 30,0-30,5 / 19,2-19,7 GHz, is divided not by 4 but 8 litters (125 MHz each) with polarization (circular - left L and right R). The integrated frequency resource will be 2500 MHz (2 beams to 1 Liter, 6 - 2 letters and 2 - to 3 letters). Ten beams with an angular size 0.75° were obtained.

3 Rating bandwidth

To assess the throughput (C) of the forward and reverse subscriber channels their energy potentials EP and the threshold ratio of received binary symbols energy E to the power spectral density of the noise - $N_0 h_p^2$ were used [7, p.48]:

$$C = \frac{EP}{h_p^2} \tag{1}$$

Since it was important to determine the throughput of channels not only in clear weather, but in the rain, the calculations of the attenuation of radiosignals with circular polarization in the rain (0.1 percent of time) for Ust-Kamenogorsk at frequencies 20 and 30 GHz were conducted, in accordance with the procedure of the ITU [11]. Let us consider the work area of the satellite uplink beams (reverse link), namely channel subscriber terminal (ST) – a satellite transponder (SR). The main characteristic of the receiving path of SR is the quality factor $Q = G_{res} / T$, where $G_{res} = 27843 / \theta^2 = 27843 / 0.75^2 = 49498.7$ (46.9 dB) - the gain of the receiving

antenna of SR and T - noise temperature of the receiving path of SR which according to [7, p.118] can be taken as 1000 K (30 dBK). Thus, the work area will be characterized by quality factor $Q = 46.9 - 30 = 16.9$ dB / K. To estimate the throughput of the channel it is necessary to consider its energy budget (Table 2).

Let us define the throughput C of the channel, provided, that the probability of erroneous reception is not more than 10⁻⁷. Table 2 shows that in this channel at the same time frequency and energy resources are limited - output power of 1 W and bandwidth $\Delta f = 125$ MHz. Therefore, a joint performance of two inequalities is required:

$$C \leq C1 = \frac{Pc}{(N_0 \times h_p^2)}, \tag{2}$$

[7, p.117]

$$C2 = C \leq \frac{\Delta fs \times m \times r}{(1 + \alpha)}, \tag{3}$$

[12, p.96]

where h_p^2 - threshold ratio of energy of received binary symbols E to noise power spectral density N_0 ; Δfs - bandwidth occupied by the transmitted signal; m - spectral efficiency modulation method; r - code relative velocity; α - rounding figure. When $\alpha = 0.25$ formula (3) takes the form:

$$C \leq C2 = 0.8 \times \Delta fs \times m \times r. \tag{4}$$

TABLE 2 Energy budget channel ST – SR

Parameters	Symbol	Value
Angular size of the beam, deg		0.75
Frequency range, GHz		30(Ka)
Output power of the ST, W		1
Diameter antenna of the ST, m		0.6
Gain of a transmitting antenna of the ST, dB		43.3
Signal losses on the transmission side, dB		1
E.I.R.P channel, dBW		42.3
Channel frequency band, MHz	Δf	125
Loss of the antenna pointing, dB		2.5
Free-space transmission loss, dB		213.8
Signal loss in the atmosphere gases, dB		0.4
Gain of a receiving antenna of the SR, dB		46.9
Noise temperature of a receiving path of the SR, K		1000
Transmission loss in rain (time percent 0.1), dB		7.9
Power of the desired signal at the receiver input of SR in clear weather, dBW	P_c	
Noise power spectral density, dBW/Hz	N_0	-127.5
Energy potential of a radio channel ST–SR, dBHz	EP	-198.6

Note: ST – subscribe (user) terminal, SR – satellite transponder

The throughput of the channel is equal to $C = \min(C1, C2)$. When $C = C1$ the throughput is determined by the energy capabilities of ST and the bandwidth is used partially. When $C = C2$ throughput of the channel is limited by frequency resource. Calculations show that with the initial data (Table 2) in clear weather, and when using signal-code structure (SCS) in the standard DVB-S2 [QPSK ($m = 2$), FEC ($r = 9/10$), $\alpha = 0.25$, $h_p^2 = 3.89$ dB] throughput of reverse channel will be equal 5.26 Mbit / s. In this case, the required frequency resource in accordance with (4) will be 3.65 MHz. The resulting throughput value corresponds approximately to the project parameters of Inmarsat-5, where it is established maximum speed of 5 Mbit / s for the reverse channel having an antenna diameter of 0.6 m.

Table 2 shows that the signal loss in the rain (0.1 percent of time) amounts 7.9 dB. In such weather conditions, the channel throughput is significantly reduced (to a value of 853 kbit / s). Transition in the rain on a new option (QPSK, $r = 1/4$, $\alpha = 0.25$, $h_p^2 = 0.75$ dB) will increase this value approximately twice (up to 1.76 Mbit / s). Usually the loading of reverse channel is relatively low, which makes it possible to use unclaimed reverse channel resource for organization of video surveillance systems [4, p.51-52].

Let us consider the work area of the satellite downlink beams, namely channel SR - ST. The initial data for calculation of energy potential of the radio channel are given in Table 3.

The calculations show that with the initial data (Table 3) in clear weather, and when using the SCS in DVB-S2 standard [8 -PSK ($m = 3$), FEC ($r = 8/9$), $\alpha = 0.25$, $h_p^2 = 6.46$ dB] the throughput of forward channel will be 224 Mbit / s. Wherein the required frequency resource will be 105 MHz. Table 3 shows that the signal losses in the rain (0.1 percent of the time) amount 3.4 dB. In such weather conditions, the channel throughput is reduced to a value

of 102 Mbit / s. Transition in the rain on a new option (8 - PSK, $r = 3/5$, $\alpha = 0.25$, $h_p^2 = 3.00$ dB) could restore its original value of throughput.

TABLE 3 Energy budget channel SR – ST

Parameters	Symbol	Value
Angular size of the beam, deg		0.75
Frequency range, GHz		20(Ka)
E.I.R.P channel, dBW		60.16
Channel frequency band, MHz	Δf	125
Loss of the antenna pointing, dB		2
Free-space transmission loss, dB		210.2
Signal loss in the atmosphere gases, dB		0.4
Gain of a receiving antenna of the ST, dB		39.8
Noise temperature of a receiving path of the ST, K		400
Transmission loss in rain (time percent 0.1), dB		3.4
Power of the desired signal at the receiver input of SR in clear weather, dBW	P_c	-112.64
Noise power spectral density, dBW/Hz	N_0	-202.6
Energy potential of a radio channel SR–ST, dBHz	EP	89.961

The total throughput of forward user channels in clear weather will be $C_{\Sigma} = (2*224 + 6*448 + 2*672) = 4480$ Mbit/s. The specific rate determines the number of subscribers, who can be connected to the network. If to focus on the specific rate adopted for cable networks, such as 30 kbit/s, it is possible to connect approximately 149 thousand subscribers. If to provide a service similar to Wild-Blue and available in the U.S. today, it is possible to connect the 44.8 thousands subscribers [4, p.1].

The total throughput could be increased by increasing of ST EIRP, but it may be that the frequency resource is not enough. Analysis of the formulas (2) and (4) shows that $C1 = C2 = 265.5$ Mbit / s by transponder EIRP 60.9 dBW. In this case total throughput is $C_{\Sigma} = 5310$ Mbit / s.

It is also necessary to verify compliance with the rules on the allowable values of power flux density (PFD) of ST signals at the earth's surface. For the usable frequency range of ST transmitter the limit of PFD reference bandwidth of 1 MHz and in a range of elevation angles $\epsilon = 5^\circ - 25^\circ$ are defined by the formula [12, p.131]:

$$PFD_{allowable} = -115 + 0.5(\epsilon - 5), \text{ dB (W/m}^2\text{)}, \quad (5)$$

For the example (Table 3) the elevation angle is $\epsilon = 23.8^\circ$ and $PFD_{allowable} = -105.6$ dB (W / m^2), and the actual value of PFD allowable is -125.1 dB (W / m^2). As the actual value of PFD significantly below the permissible level, the norms will be implemented for the other elevation angles.

4 Conclusions

10 beams with an angular size of 0.75° are determined in the practical creating of multi-beam satellite operating area in the territory of Kazakhstan.

The selection of beam size eliminates the need of direction finders used for holding the aiming point and, at the same time, provides a gain of approximately 10.5 dB

in radio line power for subscriber station as compared with a single-beam coverage area.

The multi-beam coverage area permits to increase the efficiency of spectrum using considering the uneven population density in Kazakhstan territory.

The application of adaptive coding and modulation (ACM) allows to reach a maximum bit rate in all weather conditions.

The calculation results of reverse subscriber channels throughput have shown that by choosing of signal-code structure it is possible to provide amounts of 5-6 Mb/s, which corresponds to, for example, design parameters of Inmarsat-5 satellites [1, p.1].

The total throughput of forward subscriber channels amounts 5310 Mb/s in condition of full use of declared power data (EIRP = 60.9 dBW) and frequency resource (125 MHz).

References

[1] Anpilogov B 2011 Satellite queuing system in the Ka – band *Technology and Communications, Special Issue “Sputnikovaya svyaz i veshhanie”* 16 – 21 (in Russian) <http://www.media-publisher.ru/pdf/Nom-2-2013-sait.pdf>

[2] Anpilogov B 2011 Supplements to the article "Satellite queuing system in the Ka-band *Technology and Communications, Special Issue “Sputnikovaya svyaz i veshhanie”* 96 - 98(in Russian)

[3] Testoedov N 2011 Plans of the Russian satellite constellation in the Ka-band frequencies *Technology and Communications, Special Issue “Sputnikovaya svyaz i veshhanie”* 12 - 13(in Russian)

[4] Cisar L 2011 Ka Band world experience of 2010 *Technology and Communications, Special Issue “Sputnikovaya svyaz i veshhanie”* 50 – 52 (in Russian)

[5] Afonin A Gavrilov K Satellite Networks mass of broadband Internet access in the Ka - band: trends, analysis solutions *Bulletin of the Russian Academy of Natural Sciences* No 4 4 - 9(in Russian)

[6] Babintsev A, Losev A 2011 Finding the optimal parameters multibeam satellite system *Trudy NIIR* No 2 38 - 51(in Russian)

[7] *Sputnikovye seti svyazi: Textbook* 2010 ed Kamnev V E et al Moscow: Voennii Parad 608 p(in Russian)

[8] *The number and density of population in Kazakhstan* <http://www.stat.kz/publishing/Documents/население.pdf>

[9] Broadband satellite communications in Ka - band: a systematic approach and solutions 2012 *Presentation of the company Astrium ITU Regional Workshop Almaty* 5-7 September 2012

[10] *Applications of the Radio Regulations* 2004 Volume 2 Part 2. – Geneva

[11] Recommendation ITU-R P.618-10 Propagation data and prediction methods required for the design of communication systems Earth – space <http://www.itu.int/rec/R-REC-P.618/en>

[12] *Elektromagnitnaya sovmestimost sistem sputnikovoj svyazi* 2009 ed LY Cantor and V Nozdrin Moscow: NIIR 280 p (in Russian)

Authors	
	<p>Altay Aitmagambetov, born in July 4, 1948, Almaty, Kazakhstan</p> <p>Current position, grades: Professor of Computer science and Telecommunications Department, Candidate of technical science University studies: Moscow Technical University of Communication and Information Scientific interest: satellite communications, mobile communications, spectrum monitoring Publications: more than 100 Experience: Head of radio communications department in Almaty university of power engineering and telecommunications (1981-2007), General Director of Republican State Telecommunication Inspectorate (1995-1999), Member of Director’s Board of JSC “Transtelecom” and JSC “Nursat”</p>
	<p>Yuri Butuzov, born in November 1, 1937, Irkutsk region, Russia</p> <p>Current position, grades: Professor of radio engineering department, Almaty university of power engineering and telecommunications, Candidate of technical sciences. University studies: Tomsk polytechnical institute Scientific interest: wireless communications, satellite communications Publications: more than 80 Experience: radio engineering enterprises, scientific-research institute, universities</p>
	<p>Sypatay Torekhan, 29 September 1992, Almaty, Kazakhstan</p> <p>Current position, grades: bachelor studies University studies: International IT University Scientific interest: satellite communication systems, mobile communications, adaptive arrays Publications: 2</p>

Development of multidimensional model of data for information and analytical decision-making support system

R Uskenbayeva, B Kurmangaliyeva, N Mukazhanov*

Department of Computer Science and Software Engineering, International University of Information Technologies, 34 «A»/8 «A» Manas Str./Zhandosov Str., Almaty, Kazakhstan

Received 1 February 2014, www.tsi.lv

Abstract

Today almost in all spheres of industry it is observed the urgent need of information technologies development and increase of using opportunities. This paper is devoted to creation of multidimensional model of data for information and analytical system of decision-making support for management. The information and analytical system is developed by means of OLAP technology (On-Line Analytical Processing) which in turn analyses collected materials at a real time based on strategic level of solution. In information systems this technology displays data in the form of a multidimensional cube, and necessary data for users will be expressed in the form of cube slices. The multidimensional model of data is developed for information system intended for decision-making support of higher education institutions management.

Keywords: Multidimensional model of data, data warehouse, hypercube, dimension, higher education

1 Introduction

Today for higher education institutes there is a very small amount of decision-making support systems. Recently introduced education systems and forms require a new mode of management. Higher education quality management and control is one of the main goals of HEIs. In many cases HEIs management make decisions on base of old management systems, their own practice or international practice. The information and analytical system helps university management in decision making on base of real data.

Education quality control is identified by means of quantitative and qualitative indication measures. What does the indication of education quality mean? How it can be controlled? Among academic community there are various kinds of opinions on this issue. For instance, at educational institutes quality of education is evaluated on base of students' academic progress of their quantitative and qualitative indicators, while quality of graduates' skills can be assessed by labour market demand. Employers evaluate quality of higher education institutes on base of graduate's efficiency and their contribution into a company development.

Quantitative and qualitative indicators of HEIs are as follows: faculty staff, research activity, international cooperation, alumni and their achievements, financing, national and international rankings and others.

In order to control at a real time period of higher education quality on base of modern informational technologies it is necessary to gather the previously mentioned indicators and to organize work on this big

data; to prepare and to process data for analysis; to make new methods of data modelling and analysis; big data processing with constant updating and information improving for its multidimensional analysis; on base of obtained analysis it is necessary to set up the information and analytical system for HEIs development and control of decision making data.

There are a variety of informational systems used at HEIs of our country. These systems are intended for students registering and control of educational process. Where such components like HEIs' activities data analysis, prognostic analysis and process control at a real time are not considered at nowadays existing analytical systems. Management of the decision-making support systems is based on analytical data handling and data model is aimed to use the collected data as an educational resource. The explicit measure indicators, analysis, methods and conducted monitoring used in the system must contain the whole process (working procedure) of the educational organization, and derived results must be continually analysed and compared with the targeted objectives approved by the educational institute. On its basis should be conducted error corrections and can be defined in advance HEIs' future activities. Use of new informational technologies and methods in the system conducting monitoring for the management of institutions of higher education contributes to the effective management. The OLAP technology is used to implement the system. This technology is designed for big data of indicators analysis. The main advantage of it is its capability of data analysis at a real time and quick execution of multidimensional inquiries. The OLAP

* Corresponding author - e-mail: mukazhanovn@gmail.com

technology is used in many information and analytical systems and is embed as particular service in extensively used management systems of database. As the OLAP technology works with multidimensional data warehouse, appears a task of multidimensional model set up [1, 2].

2 The multidimensional data model

The theory of acquisition is used in the multidimensional data model and presentation of aggregation indicators. On the basis of OLAP technology that presents the information in a type of cube lays the idea of the multidimensional data model. Logic of human is many-sided. When a person asks a question he sets a certain limitation, so he changes one question in many dimensions, that's why multidimensional mode of analytical process is very similar to human mind's versatility. We can say that the multidimensional data model discussed in this research work is appeared from versatility of human logic. Creation of the multidimensional data model intended for analytical system presented in this paper is composed of the following steps:

- To analyse the object sphere;
- To determine measures;
- To determine the quantitative indicators;
- To create logical model of multidimensional database. In this step we get relational schema of data warehouse in type of "snowflake" by connecting the table of facts and measure tables through logical connective [2].

Measure dimensions in the multidimensional data model are given as cube edge axis. Multidimensional cube is a multidimensional structure; multidimensional structure is a database that is closely connected with each other. Concerned multidimensional data model includes such basic concepts as hypercube, dimensions, measure indicators (members), cells, quantitative measures (indicators).

Cube consists of several dimensions, and every dimension is composed of measure indicators that form on edge of cube (figure 2 and figure 3). Cells can be composed of quantitative indicators or can be empty. Inspection of data saved in cube is conducted as a slice set composed of selected measures and quantitative measures (indicators).

Dimensions – a set of one-type or a polytypic object that defines a content of quantitative indicators. It can be simple one-dimensional or multileveled hierarchic structural by organization structure.

Measure indicators (members) – a table row or a table column consisted of indicators summary. The summary of inner objects is measure components (members). For instance, the components of an academic degree: baccalaureate, magistrate, Ph.D. The measure components can be various typed as: qualitative indicators of students, indicators of research works, time indicators: academic year, term, and week. The

components of every measure must be organized structural correctly and conveniently for expert. If measure indicators are organized by hierarchic structure, one typed objects will be placed on the same level. It is accepted to indicate measure indicators in a type of multidimensional cube's edge (axis).

Quantitative measure (fact) – a size in a quantitative type and the main component of analysis. There can be one or several quantitative indicators in an OLAP cube. For example, numerical strength of students and professional and teaching staff, finance and the quantitative measures of research works.

Cells – a cube structure that conforms to exact value of measures. During demonstration, cells are placed inside of cube and it is accepted to indicate quantitative measure value according to these cells.

We take the following marks through mentioned detection in accord with the theory of acquisition:

D – a set of dimensions, d_i - element of a set D that is $D = \{d_1, d_2, \dots, d_n\}$.

n – a number of dimensions;

$M_{d_i} = \{m_{1_i}, m_{2_i}, \dots, m_{k_i}\}$ – a set of measure indicators (dimensions' elements) or subset of a dimensions, k – a number of derived inner elements set, $i = 1 \dots n$ an amount of marks in dimension.

M – an amount of cube measure indicators (dimensions' elements):

$$M = M_{d_1} \cup M_{d_2} \cup \dots \cup M_{d_n} \quad [4, 5, 6]$$

Definition 1: A set M is a subset of a set D if every element of M is also an element of D :

$$M \subset D := m_{1_i} \in M \Rightarrow m_{1_i} \in D, \quad (1)$$

In this case, M is a subset of D , D - M superset. If $M \subset D$ and $M \neq D$, we call M – a proper subset of D (figure 1). Any other subset of D contains at least one element of D , but not all of its elements. Note that the empty set is a subset of every set. $\emptyset \subset M$ for every set M [16].

Definition 2: Two sets are identical if and only if they have exactly the same members. So $M = D$ if for every $m_{1_i}, m_{1_i} \in M \Leftrightarrow m_{1_i} \in D$.

$$M = D := M \subseteq M \ \& \ D \subseteq M, \quad (2)$$

In any sense of having a problem are usually considered a subset of "most" of the set U , which is called the universal set. In our problems, the universal set is a set of cubes. For example, Figure 1 shows a universal set U and two subsets - of M and D , $M \subset D$. Figure 1 is called the Euler-Venn's diagram [16].

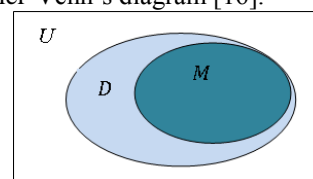


FIGURE 1 Euler-Venn's diagram

If we consider a cube as a set of cells, then hypercube – $H(D, M)$, D and M – accordingly they are inner summaries of cube. If we mark cube cell as h , then $h \in H$, and measure indicators also will be given as cells M_h , hereby: $M_h \in M$ [5].

As mentioned above, user can see the data as a cut set composed of several measures. If we mark a summary of measures selected for cut set as D , if we mark a summary of measure indicators relating to selected measures as M' , then a cut set of cube will be indicated as $H'(D, M')$. Dimensions of cut set – $d_i \in D'$ and measures $m_{k_i} \in M'$. Slice set dimensions are formed by consolidation of measure summary $M_{d_1} \cup M_{d_2} = M'$, and quantitative measures are formed by intersection of measure indicators' summary (an aggregative value given in an intersection axis of dimensions): $c_i = \{c_{1,1}\{m_{1_1}, m_{1_2}\}, c_{1,2}\{m_{1_1}, m_{2_2}\}, \dots, c_{k_1,1}\{m_{k_1}, m_{1_2}\}, c_{k_1,2}\{m_{k_1}, m_{2_2}\}, \dots, c_{k_1-1,k_2-1}\{m_{k_1-1}, m_{k_2-1}\}, c_{k_1,k_2}\{m_{k_1}, m_{k_2}\}\}$

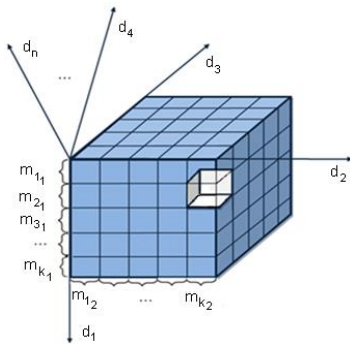


FIGURE 2 Multidimensional data model given as a cube

If a cube consists of n - dimensions, then it will be able to determine a quantity of all possible two-dimensional slice sets by the following formula:

$$N_s = \frac{n^2 - n}{2}, \tag{3}$$

Here, N_s - a quantity of two-dimensional slice sets outgoing from cube;
 n - a number of dimensions.

A set of quantitative measures is marked as x_{k_i} , quantitative measures of dimensions x_{k_i} ($i=1..n, k=1..|d_i|$) - values of measures are given as aggregative summand. Usually aggregative summand is determined by horizontal and vertical calculating of quantitative measures situated in intersection axis of dimension inner summaries (measures).

Bringing higher educational institutions' measures through cube dimensions to the marks derived by the theory of acquisition:

Set of dimensions: $D = \{\text{academic years, levels of educational system, specialties, education forms,}$

$\text{indicators of students, international students, student clubs, graduates, professional and teaching staff, academic degree, academic title, research activities, scientific infrastructure, international relations, selection committee, prizes for undergraduate students, financial and educational programs, quality evaluation, accreditation, international rating, accreditation organizations, infrastructure, sport buildings, institutions, departments}\}$;

Set of dimensions' inner measures: $M_{d_i} = \{\{2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, \dots\}, \{\text{Total amount, amount of governmental grants, ..., amount of international students}\}, \{\text{Computer science, Information systems, Software engineering, ...}\}, \dots\}$

Set of dimensions inner measures:
 $m_{1_1, \dots, k_1} = \{2009-2010, 2010-2011, 2011-2012, 2012-2013, 2013-2014, \dots\}$;

$m_{1_2, \dots, k_2} = \{\text{Total amount, amount of governmental grants, ..., amount of international students}\}$;

$m_{1_3, \dots, k_3} = \{\text{Computer science, Information systems, Software engineering, ...}\}$

...
 $m_{1_4, \dots, k_4} = \{\dots, \dots, \dots\}$

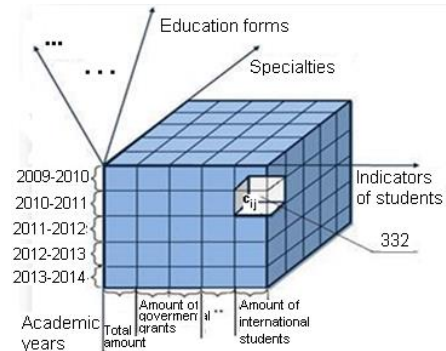


FIGURE 3 Higher educational institutions' dimension measures are given in the shape of a multidimensional cube

Hierarchical structure. Hierarchic structure is embedding to data model. It allows passing to hierarchic level at any time. L – if we will consider d_1 dimension composed of hierarchic levels, e.g. d_1 – time dimension, an academic year – term – a week. In the hierarchic structure, initial values are situated in lower level (FIGURE 4).

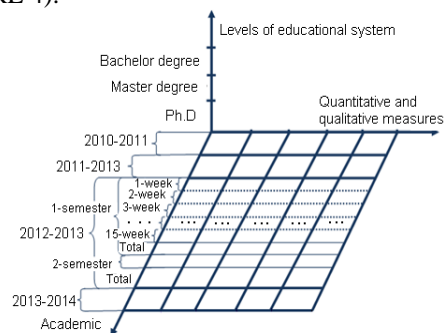


FIGURE 4 Hierarchic structure of an academic year

In this structure quantitative measures are taken from the lowest level and its' value is passed on higher levels. In order to conduct quickly the process of passing on hierarchic levels, calculating values of various levels must be saved as multidimensional model.

3 Data warehousing logical design

There is a proper order of creation and connection tables of data warehouse. There two types of tables: fact table and dimensions table. All the information of informational sphere components in resolving this task must be included in a logical data model creation process. They are intended to create entity-relationship logic database model, also they provide with presentation of data structure in defining the objects. The main motive of

selection of this notation is that it is independent from database physical creation projection and used informational platform. In the paper takes place information about information and analytical system data warehouse model in type of "snowflake" schema created by the project "Higher education quality control and corporative informational system of management" (model of data warehouse is presented in the figure 5). In computing, a snowflake schema is a logical arrangement of tables in a multidimensional database such that the entity relationship diagram resembles a snowflake in shape. The snowflake schema is represented by centralized fact tables, which are connected to multiple dimensions. "Snow flaking" is a method of normalizing the dimension tables in a star schema [15].

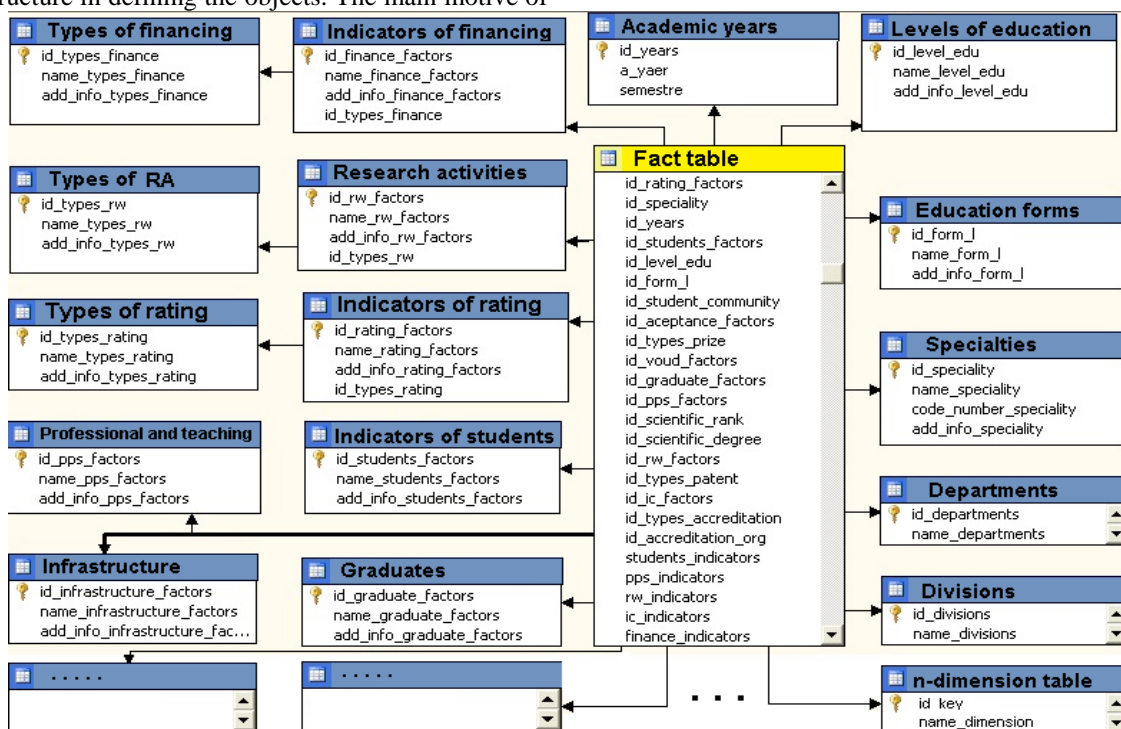


FIGURE 5 Relational schema of data warehouse in type of "snowflake"

When it is completely normalized along all the dimension tables, the resultant structure resembles a snowflake with the fact table in the middle. The principle behind snow flaking is normalization of the dimension tables by removing low cardinality attributes and forming separate tables.

Multidimensionality of an information and analytical system must be in the following three levels:

1. Multidimensional data presentation - the means of the end user providing multidimensional visualization and a manipulation with data;
2. Multidimensional processing – means of a formulation of multidimensional inquiries and the processor, able to process and execute such inquiry;

3. Multidimensional storage – the means of a physical data structure providing effective implementation of multidimensional inquiries.

The first two levels without fail are present in the system means. The third level, though is widespread, is not obligatory as data for multidimensional representation can be taken and from usual relational structures.

4 Conclusions

This article provided with the following information:

- The necessity of information and analytical system in decision making support in higher educational institutions;
- Basic definitions of multidimensional data model of one of the most important components of




information and analytical systems – hypercube structure are given;

- The theory of acquisition was used in defining the hypercube structure;
- Information and analytical system's data model meant for higher educational institution's management was prepared on the basis of hypercube.

The analytical system for higher educational institutions' management was created due to the possibilities of OLAP technologies. OLAP technologies are means intended to data handling in a certain time. The modern OLAP servers have the possibilities of quick handling of solid information taken from various sources; presenting cut sets, different diagrams and other forms of information to users. We think that for these valuable reasons the OLAP technologies are very useful for heads of higher educational institutions in decision making on basis of gathered data.

References

- [1] Uskenbayeva R K, Mukazhanov N K 2013 *VESTNIK of KazNTU*, 4(98) Almaty, Kazakhstan 106
- [2] Ermolkevich A A, Nezhurina M I 2010 Postroenie modeli informatsionno-analiticheskoi sistemy monitoring obrazovatel'nogo protsesssa prinalichiy biznes-zakazchika *Biznes-informatika* 2(12) (in Russian)
- [3] Kulagin V P, Matchin V T 2010 Matematicheskoe modelirovanie OLAP-kuba v konexite agregirovaniya prostykh Ierarkhicheskikh izmereniy *Izvestiya Tomskogo politekhnicheskogo universiteta* 5 72 (in Russian)
- [4] Kuznetsov S D, Kudryavtsev Yu A 2009 A Mathematical Model of the OLAP Cubes *Programming and Computer Software* ISSN 0361-7688 35(5) 257–265
- [5] Zabolotnev M S Metody predstavleniya informatsii v razrezhnnykh gipercubakh dannih <http://www.olap.ru/basic/theory.asp> 18 Feb. 2014 (in Russian)
- [6] Hrustalev E M 2014 Agregatsiya dannih v OLAP-cubah <http://www.olap.ru/home/mut.asp> 17 Feb. 2014 (in Russian)
- [7] Nandi A, Cong Yu, Bohannon P, Ramakrishnan R January 2012 Data Cube Materialization and Mining over MapReduce *Transactions on knowledge and data engineering* 6(1)
- [8] Mironov V V, Makarova E S 2012 Agregatsiya pokazateley v OLAP-cubepri svedeniy po zavisimym izmereniyam *VESTNIK-UGATU* T 16 3(48) 180–186 (in Russian)
- [9] Lucian B 2010 Optimized Data Indexing Algorithms for OLAP Systems *Database Systems Journal* 1(2)
- [10] Golfarelli M, Rizzi S 2009 Data Warehouse Design. McGraw-Hill *Modern Principles and Methodologies*
- [11] You J, Xi J, Zhang C 2008 A High Performance Large Scale Data Warehouse *IEEE, In International Multi-Symposium of Computer and Computational Sciences*. 200–202
- [12] Zhang D, Zhai C, Han J 2009 Topic cube: Topic modeling for olap on multidimensional text databases *Proceedings of the Ninth SIAM International Conference on Data Mining, SIAM* 1124–1135
- [13] Nedjar S, Casali A, Cicchetti R, Lakhal L 2007 Emerging Cubes for Trends Analysis in Olap databases *Lecture Notes in Computer Science*. 4654 135–144
- [14] Cuzzocrea A, Song I-Y, Davis K C 2011 Analytics over large-scale multidimensional data: the big data revolution! *In Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP, DOLAP*. New York, NY, USA, ACM. 101 – 104
- [15] Berger A B, Gorbach I V, Melomed E L, Sherbinin V A, Stepanenko V P 2007 *Microsoft SQL Server 2005 Analysis Services. OLAP i mnogomerniy analiz dannih* SPb: BHV-Peterburg 928 (in Russian)
- [16] Borschev V, Partee B September 6, 2001 Lecture 1. Basic Concepts of Set Theory *Mathematical Linguistics* p. 1

Authors	
	<p>Raissa Uskenbayeva, born in September 19, 1953, Kazakhstan</p> <p>Current position, grades: Vice-rector on the academic affairs, professor department of CSSE University studies: International University of Information Technologies</p> <p>Scientific interest: Macro- and micro-economics, finance and banking, Industrial Automation and Control Theory, Marketing, Management and logistics, Information technology and software engineering, Informatics problems</p> <p>Publications: more than 100 scientific articles, monographs on: <i>Theory of control and automation industry, Information Technology and Systems, Reliability of mathematical and software IP</i></p> <p>Experience:</p> <ul style="list-style-type: none"> • 2012- at present Almaty: International University of Information Technologies, Vice-rector on the academic affairs • 2003-2012 Almaty: Kazakh National Technical University after K. I. Satpaev, The head of the Depart. «Software of systems and nets» of the Institute Information Technologies, The Doctor of Science, professor • 1999-2003 Almaty: Doctorate at Kazakh National Technical University after K. I. Satpaev • 1987–1999 Almaty: Kazakh National Technical University after K. I. Satpaev, The Candidate of Technical Science, Docent • 1981–1987 Almaty: Kazakh National Technical University after K. I. Satpaev, tutor • 1978-1981 Almaty: Research Institute of the State Planning Committee of the KazSSR, The Junior Research Fellow • 1975–1978 Almaty: Almaty Special. Design Bureau of The Ministry of Telecommunications Industry USSR, Engineer-mathematician
	<p>Bikesh Kurmangaliyeva, born in November 03, 1963, Kazakhstan</p> <p>Current position, grades: PhD studies University studies: International University of Information Technologies</p> <p>Scientific interest: IT management, Database, Distributed data processing, Decision support systems, Artificial intelligence</p> <p>Publications: more than 10 scientific articles (IT management, Database, Distributed data processing, Decision support systems)</p> <p>Experience:</p> <ul style="list-style-type: none"> • 2012- at present Almaty, International University of Information Technologies, Ph.D student • 2008–2010 Almaty: Kazakh National Technical University after K. I. Satpaev, studied in magistracy
	<p>Nurzhan Mukazhanov, born in November 26, 1986, Kazakhstan</p> <p>Current position, grades: Ph.D studies University studies: International University of Information Technologies</p> <p>Scientific interest: Database, Information and analytical systems, Distributed data processing, Decision support systems, Artificial intelligence</p> <p>Publications: more than 10 scientific articles (Distributed data processing, Information and analytical systems)</p> <p>Experience:</p> <ul style="list-style-type: none"> • 2012- at present Almaty, International University of Information Technologies, Ph.D student • 2010–2012 Almaty: Kazakh National Technical University after K. I. Satpaev, tutor • 2008–2010 Almaty: Kazakh National Technical University after K. I. Satpaev, studied in magistracy

On tracking ability analysis of linear extended state observer for uncertain system

Shangyao Shi*, Shiping Zhao, Jun Li

Department of Measurement and Control, Sichuan University, Chengdu 610041, P. R. China

Received 1 March 2014, www.tsi.lv

Abstract

It is known that the linear active disturbance rejection control (LADRC) is very an effective approach to control the uncertain systems. The linear extended state observer (LESO) is the major link of the LADRC, so this paper presents a modified LESO, which is used to track the state variables and estimate the unknown total disturbance. Furthermore, this paper redefines the "time-scaling" of the plant, which is a function with respect to the amplitude of the unknown total disturbance. It is first time to present the specified formula of the maximum sampling-period of LESO for some existing plants. On the another hand, the tracking ability of the designed LESO is quantitatively described in this paper. The discussions and analysis, especially the quantitative formulas presented in this paper, will help the scholars and engineers to design the LESO in practice.

Keywords: LADRC, LESO, Modified LESO, "Time-Scaling", Sampling-Period

1 Introduction

The linear extended state observer (LESO) [1], as a mayor link of linear active disturbance rejection control (LADRC), was proposed by Prof. Gao in 2003. Since then, it is applied in many industrial fields, and shows promising control performance in practice. Recently, some breakthroughs of theoretical researches on LESO were obtained, such as in this paper [2], the rigorous proof of asymptotical convergence of LESO was given with some boundary constraint. And it was proven in [3] that the convergence of non-linear extended state observer (ESO) for a class of multi-input multi-output nonlinear systems with uncertainty can be feasible if the non-linear functions for observer are properly constructed. The researches in [4] showed that estimation and tracking errors of LESO are bounded, with their bounds monotonously decreasing with their respective bandwidths for large dynamic uncertainties. In particular, the observer bandwidth and closed-loop bandwidth were presented in [1], which is helpful in understanding the LESO easy from a physical perspective, and contributes to spreading out LESO applied in various fields.

In contrast to both theoretical and qualitative analysis of the LESO, this paper makes an offer to quantitatively analyses the latent tracking ability of LESO, which decided the disturbance rejection ability of the specific controller based on LADRC. This paper has presented a modified LESO, which not only track the state variables of plant, but also estimate the unknown total disturbance including the external disturbance and unknown model disturbance but excluding the known partial model, and

redefined the "time-scaling" of the plant. Moreover, it is first time to present some specified formulas to judge the tracking ability of the LESO, and present some criterion to design the LESO for some specified plants. Following the principia, described in this paper, the LESO will be properly designed soon and implement well, which is proven by the simulations in section III.

2 Analysis of the Tracking ability of the LESO

In general case, considering a n th order system, which may be nonlinear and time-varying, with a uncertain disturbance as follow:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \vdots \\ \dot{x}_n = f(X, w, t) + bu \\ y = x_1 \end{cases}, \quad (1)$$

where X is the system state variable vector $[x_1 x_2 \cdots x_n]^T$, u is the control output of the controller, and y is the measure output. $f(\cdot)$ is a uncertain generalized disturbance including the unknown system dynamics module, also named total distance including the internal disturbance and the external disturbance.

As we know that the corresponding linear extended state observer (LESO) of above system (1) with differential form is constructed as follow:

* Corresponding author- Tel: +86-28-85405310; fax: +86-28-85461088; E-mail: peter.shi@126.com

$$\begin{cases} e = y - z_1 \\ \dot{z}_1 = z_2 + \beta_1 e \\ \dot{z}_2 = z_3 + \beta_2 e \\ \vdots \\ \dot{z}_n = z_{n+1} + \beta_n e + bu \\ \dot{z}_{n+1} = \beta_{n+1} e \\ y = x_1 \end{cases}, \quad (2)$$

where the new variables $z_1, z_2, \dots,$ and z_n in (2) are the estimates of state variables $x_1, x_2, \dots,$ and x_n respectively, while the variable z_{n+1} , namely known as an augmented state variable, is the estimate of $f(\cdot)$ named the total disturbance.

A problem is proposed that how to enhance the tracking ability of the specified LESO, in other words, how much little tracking errors of the state variables can be actively got by the LESO within the constraints of the sampling-period in practice. This is very important to properly design the ADRC, because it determines the disturbance rejection ability of ADRC based on LESO.

In this section, the state variables tracking ability of LESO is discussed. Let us investigate the following plant system as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \vdots \\ \dot{x}_n = f(X, w, t) + g(X, t) + bu \\ y = x_1 \end{cases}, \quad (3)$$

where $f(X, w, t)$, the unknown total disturbance, including the external disturbance and the unknown part of the plant's model, $g(X, t)$ is the known part of the plant model. A modified LESO of the system (3) is proposed and constructed as follow:

$$\begin{cases} e = y - z_1 \\ \dot{z}_1 = z_2 + \beta_1 e \\ \dot{z}_2 = z_3 + \beta_2 e \\ \vdots \\ \dot{z}_n = z_{n+1} + \beta_n e + g(X, t) + bu \\ \dot{z}_{n+1} = \beta_{n+1} e \\ y = x_1 \end{cases}, \quad (4)$$

where z_{n+1} , extended state variable, estimates the unknown total disturbance, which including the unknown part of the plant's model. Owing to including the known part of the plant model in nth equation of expression (4), the modified LESO reduces the requirements of work speed of hardware, and improve the control precision and performance of the controller.

The conception of "time-scaling", denoted as ρ , is proposed and defined in [5]. Similarly, there the "time-scaling" of the plant (3) can be defined as bellow.

Definition 2.1: the plant's "time-scaling", denoted as ρ , of plant (3) is redefined as

$$\rho = \frac{1}{M^{\frac{1}{n}}}, \quad (5)$$

where $M = \max_{\substack{|w| < r_0 \\ |x_1| < r_1 \\ |x_2| < r_2 \\ \vdots \\ |x_n| < r_n}} f(x_1, x_2, \dots, x_n, w, t)$, and r_0, r_1, \dots, r_n are

bounded for any $0 \leq t < \infty$. And the plant with $\rho = 1$ is named typically "canonical system" or "canonical plant".

The following theorems and proofs are just discussed about the second order systems, but the conclusions are also correct for the nth order system.

Theorem 2.1: The maximum sampling-period, denoted as τ_{max} within the permit of the tracking estimation error ε demanded in practice, of the LADRC based controller is depended on the plant's inherent "time-scaling", defined as ρ in definition 2.1.

Proof: A typical second order plant, represented by the state space equation, can be described as:

$$\begin{cases} \dot{X}(t) = AX(t) + B(u(t) + g(X, t)) + E\dot{f}(t) \\ y(t) = CX(t) \end{cases}, \quad (6)$$

where $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $E = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $C = [1 \ 0 \ 0]$,

$X(t) = [x_1(t) \ x_2(t) \ x_3(t)]^T$ and $\dot{X}(t) = [\dot{x}_1(t) \ \dot{x}_2(t) \ \dot{x}_3(t)]^T$ is the state vector and its derivative, respectively. And $g(X, t)$ is the known part of the plant's model, and $\dot{f}(t)$ is the derivative of $f(X, w, t)$, the uncertain function, which is bounded and continuous or piecewise continuous during an interval $0 \leq t \leq T$ for any $T > 0$.

The LESO of the above plant (6), according to preceding description, is constructed as

$$\begin{cases} \dot{Z}(t) = AZ(t) + B(u(t) + g(X, t)) + L(y(t) - \hat{y}(t)) \\ \hat{y}(t) = CZ(t) \end{cases}, \quad (7)$$

where $L = [3\omega_0 \ 3\omega_0^2 \ \omega_0^3]^T$, is the observer gain vector, $Z = [z_1 \ z_2 \ z_3]^T$ is the estimate vector of X.

Then, subtracting (6) from (7) on both sides at the same time, the dynamics estimation error is obtained as:

$$d\dot{E}(t) = (A - LC)dE(t) + E\dot{f}(t), \quad (8)$$

where $dE(t) = [\varepsilon_1 \ \varepsilon_2 \ \varepsilon_3]^T$, $d\dot{E}(t) = [\dot{\varepsilon}_1 \ \dot{\varepsilon}_2 \ \dot{\varepsilon}_3]^T$, and $\varepsilon_i = x_i(t) - z_i(t), i = 1, 2, 3$.

Assume the sampling-period is τ , the discrete form of (8) becomes as:

$$dE(k) = (I + (A - LC)\tau)dE(k-1) + E\dot{f}(k-1) = \left(I + ((A - LC)\tau)^k dE(0) + \sum_{i=1}^k (I + (A - LC)\tau)^{k-i} E\dot{f}(k-1)\tau \right), \quad (9)$$

$$= \left(I + ((A - LC)\tau)^k dE(0) + \sum H^{k-i} E\dot{f}(k-1)\tau \right)$$

where $(I + ((A - LC)\tau) = \begin{bmatrix} 1 - 3\omega_0\tau & \tau & 0 \\ -3\omega_0^2\tau & 1 & \tau \\ -\omega_0^3\tau & 0 & 1 \end{bmatrix}$ and

$$H^i = (I + ((A - LC)\tau)^i E, i = 1, 2, \dots, k.$$

Assuming that $z_i(0) = x_i(0), i = 1, 2, 3$, i.e. those initial values are equal, at the beginning of the LADRC based controller acting.

Let $\omega_0 = m/\tau$, where $0 < m < 1$ is proposed [6], then the vector term H^i can be computed as:

$$H^i = \begin{bmatrix} \frac{\tau^2}{2} i(i-1)(1-m)^{i-2} \\ \tau i((i-2)m+1)(1-m)^{i-2} \\ ((i-1)(i-2)m^2 + 2(i-2)m+2) \frac{(1-m)^{i-2}}{2} \end{bmatrix}. \quad (10)$$

Having an insight into (10) and (9), it is obvious that the tracking estimation error, $\varepsilon_1 = x_1 - z_1$, is obtained as follow:

$$\begin{aligned} \varepsilon_1 &= \sum_{i=1}^k H^{k-i} E\dot{f}(k-1)\tau \\ &= \sum_{i=1}^k \frac{\tau^3 (k-i-1)(1-m)^{k-i-2} (k-i)}{2} \dot{f}(k-1) \\ &= \sum_{i=0}^{k-1} \frac{\tau^3 (i-1)(1-m)^{i-2} i}{2} \dot{f}(k-1) \\ &\leq \frac{\tau^2}{m^3} (f(k) - f(k-1)) \\ &\leq \frac{M\tau^2}{m^3} \end{aligned} \quad (11)$$

So as long as the design requirements error ε satisfies $\frac{M\tau^2}{m^3} \leq \varepsilon$ in practical engineering, then the actual tracking estimation error ε_1 will meet the design requirement. As a result of the above analysis, the following inequality expression is yielded as

$$\tau \leq \sqrt{\frac{\varepsilon m^3}{M}} = \sqrt{\varepsilon m^3} \rho, \quad (12)$$

which can be transformed into

$$\tau \leq \tau_{\max}, \quad (13)$$

where $\tau_{\max} = \sqrt{\varepsilon m^3} \rho$. That is to say the ‘‘time-scaling’’ of the specific plant demands that the sampling-period of controller must be smaller than the τ_{\max} , determined by (13), to meet the desired design allowable error ε .

Remark 2.1: Those plants, whose ‘‘time-scaling’’ are larger than the minimum ‘‘time-scaling’’ ρ_{\min} which the controllers possess, can be controlled well by these LADRC based controllers.

Proof: The proof is obvious, form (12), the following inequality expression can be derived $\rho \geq \frac{\tau}{\sqrt{\varepsilon m^3}}$, which can also be converted as:

$$\rho \geq \rho_{\min}, \quad (14)$$

where $\rho_{\min} = \frac{\tau}{\sqrt{\varepsilon m^3}}$, which is decided by the sampling-period and the allowable tracking error of the controller. That is to say, the specific sampling-period τ and the demanded the tracking estimation error of the controller is given, then the disturbance rejection ability of this LADRC based controller is determined. In other words, the maximum amplitude of the plant’s unknown disturbance is $\frac{1}{\rho_{\min}^2}$ for second plant.

3 Simulations

Assume there is a plant, which is a second-order system and described in state space as follow:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x_1, x_2, t) - 100.4x_2 + 230u, \\ y = x_1 \end{cases} \quad (15)$$

where $f(x_1, x_2, t)$ is the unknown disturbance excluding the known part of the plant’s model. From previous analysis, it is known that its corresponding modified LESO is:

$$\begin{cases} e = y - z_1 \\ \dot{z}_1 = z_2 + 3\omega_0 e \\ \dot{z}_2 = z_3 + 3\omega_0^2 e - 100.4z_2 + 230u, \\ \dot{z}_3 = \beta_3 \omega_0^3 \\ y = x_1 \end{cases} \quad (16)$$

where, we adopt the sampling-period $\tau = 1ms$, the observer bandwidth $\omega_0 = 300$ and $u = 0$ respectively. Let us investigate the performance of above specified LESO on the various plants with different ‘‘time-scaling’’. Assuming there are three plants with different unknown

disturbance $f(x_1, x_2, t)$, which are designated as the following three various expression:

A. $f(x_1, x_2, t) = 100\sin(6.28t)$,

B. $f(x_1, x_2, t) = 1000\sin(6.28t)$,

C. $f(x_1, x_2, t) = 4000\sin(6.28t)$,

which are regarded as three various plants whose “time-scaling” are $\rho_1 = 0.1$, $\rho_2 = 0.0316$ and $\rho_3 = 0.0158$ respectively. The simulation results are shown in figure 1 with the results of the plants A, B and C corresponding to subgraph (a), (b) and (c) respectively. And subgraph (d) in figure 1 shows the tracking errors along with the output y of the above three plants.

The subgraph (d) in figure 1 demonstrates that the tracking error is the largest on the plant C (the green curve in subgraph (d)) and that is the smallest on the plant A (the red curve in subgraph (d)). That to say, the larger the “time-scaling” of plant, the smaller of the tracking error is for the same LESO.

Suppose that the demanded tracking estimate error, in practice, is less than 0.02, then the above LESO can well track those plants whose “time-scaling” are larger than 0.043, i.e. $\rho \geq \rho_{\min} \approx 0.043$, which can be obtained from inequality expression (14).

On the other hand, figure 2 shows the simulation results about the LESO with different sampling-periods acting on the same plant A expressed above. The sampling-period of LESO, shown in figure 2, is $\tau_1 = 1ms$ in subgraph (a), $\tau_2 = 2ms$ in subgraph (b), $\tau_3 = 4ms$ in subgraph (c), and the three tracking errors are displayed

in subgraph (d) in figure 2. The subgraph (d) in figure 2 demonstrates that the tracking error is the smallest with sampling-period $\tau_1 = 1ms$ (the red curve in subgraph (d)) and in contrast, that is the largest with sampling-period $\tau_3 = 4ms$ (the green curve in subgraph (d)). In other words, the smaller the sampling-period of LESO is, the higher the tracking precision is without considering the sampling noise. Assuming the required tracking error is $\varepsilon \leq 0.02$, this demands the maximum sampling-period of LESO is $\tau_{\max} \approx 2.3ms$ obtained from the inequality expression (13) to track the state variables well.

4 Conclusions

In this paper, the tracking performance of the LESO is discussed by the quantitative description. It is first time that the maximum sampling-period of LESO is presented by a specified formula (13) for an existing plant. On the other hand, the tracking ability of the designed LESO is presented by formula (14). The theory analysis in section II and simulation results in section III both show that the smaller the plant’s “time-scaling”, the larger tracking error is, using the same LESO. Conversely, the shorter of the sampling-period of LESO, the higher the tracking accuracy is for the same plant. The above discussions and analysis, especially the quantitative formulas (13) and (14) will help the scholars and engineers to design the proper LESO in practice.

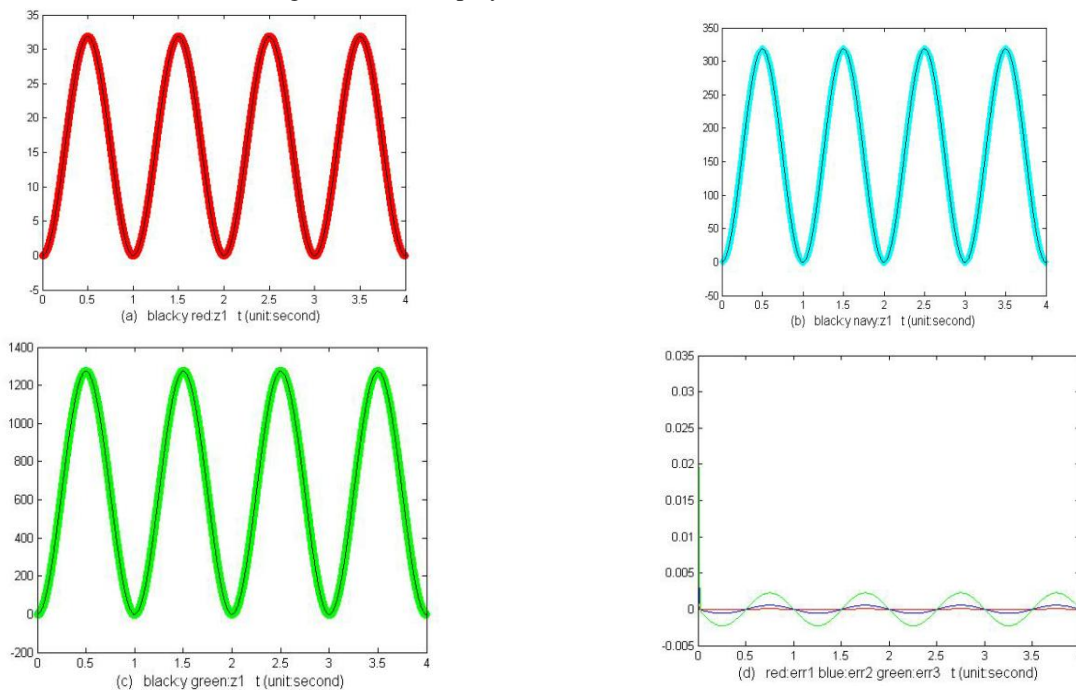


FIGURE 1 The comparison of tracking errors of the same LESO acting on various plants with different “time -scaling”

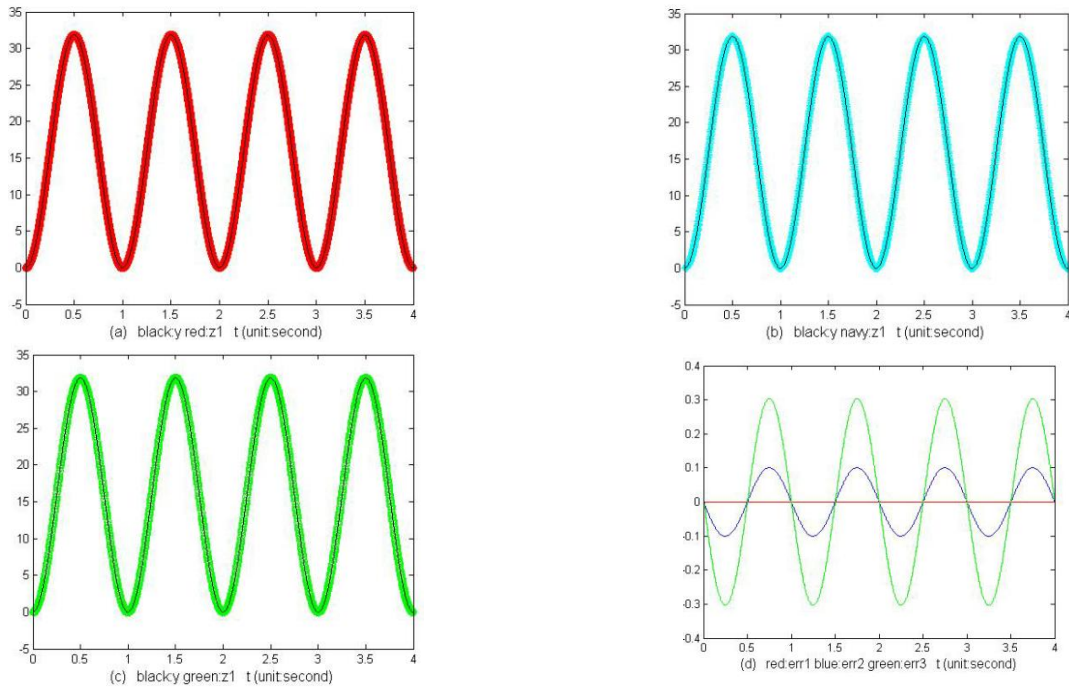


FIGURE 2 The comparison of tracking errors of the LESO with different sampling-period acting on the same plant

References

[1] Gao Z 2003 Scaling and Bandwidth-Parameterization Based Controller Tuning *In Proc. of the 2003 American Control Conference, Denver, CO, USA, June 2003* 4989-4996




[2] Bao Zhu Guo, Zhi Liang Zhao 2011 On the convergence of an extended state observer for nonlinear systems with uncertainty. *Systems & Control Letters* **60** 420-430

[3] Guo B Z, Zhao Z L 2012 On convergence of non-linear extended state observer for multi-input multi-output systems with uncertainty *IET Control Theory and Applications* **6**(15) 2375-2386

[4] Wenchao Xue, Yi Huang 2013 On Frequency-domain Analysis of ADRC for Uncertain System *In Proc. of 2013 American Control Conference, Washington, DC, USA, June 17-19, 2013*

[5] Han J 1994 The time scale and the nonlinear PID controller *In Proc. of the 1994 Chinese Control Conference, Taiyuan, China, August 15-22, 1994*

[6] Yoo D 2007 Optimal fast tracking observer bandwidth of the linear extended state observer *International Journal of Control* **80**(1) 102-111

Authors	
	<p>Shangyao Shi</p> <p>Current position, grades: Chief Engineer in Wuhan HuaGong Laser Engineering corporation, China.</p> <p>University studies: received his Master of Engineering in Manufacture of Machinery and Automation (1997) and now is studying at Sichuan University with a major of Measurement and Control Technology and Instrumentation Program for PhD (2014).</p> <p>Scientific interest: current research interests include Servo Control, Embedded System, Artificial Intelligence and Distributed Control Systems.</p>
	<p>Shiping Zhao</p> <p>Current position, grades: full professor of manufacturing science and engineering Department, Sichuan University.</p> <p>University studies: received his PhD in engineering precision instruments and machinery (1991) from Chongqing University.</p> <p>Scientific interest: He makes a great deal research on robot technology and non-destructive test.</p>
	<p>Jun Li</p> <p>University studies: M.Sc. in Engineering (2008) from southwest University of science and technology. Now he is studying at Sichuan University with a major of Mechatronic Engineering Program for PhD (2014).</p> <p>Scientific interest: the Electrical discharge machining, Micro Electrical discharge machining and Artificial intelligence used in the mechanical engineering.</p>

Using genetic algorithms on service facilities layouts design

Wei Shang¹, Maozhu Jin²*, Jie Wang²

¹ Chengdu University of Information Technology, Southwest Airport Economic Development Zone no.24, Chengdu 610225, China

² Business School, Sichuan University, Wangjiang Road.29, Chengdu 610064, China

Received 1 March 2014, www.tsi.lv

Abstract

Traditionally, the objective of a service facility layout problem is to minimize the material handling cost and optimize the physical spaces that support a productive activity of the service system. Service delivering interface has direct impact to the efficiency and effectiveness of service system in light of customer's contact. Layout problem of service facilities has received little attention in the literature. Due to complexity and combination of the facility layout problems, the genetic algorithm is put forward that codifies a chromosome and uses a direct representation to assign the departments to the different branch areas, enables good solutions to be justified in real instances of the problem in tourist service centre. The solution is shown on a decision plan representing the graphic layout of the departments within the service facility. This application is highly flexible that the decision maker can accordingly define the departments' number to be accommodated, facility sizes, segment sizes, and algorithm parameters to be added or simplified.

Keywords: Layout design; Service facility; Genetic algorithm; Tourist service centre

1 Introduction

When designing a service facility, one of many important issues is the need to minimize material handling costs and optimize the physical spaces that support a productive activity of the service system. A good facility layout can reduce these costs by at least 10%-30% and 8% of the U.S. gross national product has been spent on new facilities every year since 1955 and modifications of existing facilities are not included [1]. To find the optimization of the design of facility layout is important in both manufacturing and service industries [2]. Closely related methods can be used when designing of airports, public service platforms and similar installation where efficient flow of customers, working staff, supporting materials and information is of great importance.

The Facility Layout Problem is based on the optimal allocation of all the departments that are involved in a productive process in order to achieve the economy of resources such as space and time. A facility or department is a physical entity that facilitates the development of any productive process. The minimum parameters required to solve the problem are: number of departments, size of the office, size of the departments to be located, flow matrix between the departments of the branch and the unit transportation cost matrix. Absolute costs are not necessary in order to compare the layout alternatives, so all the unit costs are assumed to be equal (even all equal to unity) in practice [3]. A flow matrix can be changed for a perception matrix where the desirable proximity between departments is represented. As a

solution of the problem, the optimal position of each department in the branch is obtained.

In recent years, a lot of suboptimal approaches have been developed to solve Facility Layout Problems (FLPS). Given that this problem falls into the class of NP-complete problems, suboptimal methods are used to solve it. Previous works had used techniques like Tabu Search [4], Simulated Annealing [5] and Genetic Algorithms [6]. The application of Tabu Search to the Facility Layout Problem has a strong relation with the Quadratic Assignment Problem (QAP). The results suggest that a good Tabu list sizes increase with dimension of the problem and the best known solutions were obtained using less CPU time than previously reported in the literature for the standard problems tested. An adaptation of Simulated Annealing procedure based on the model to solve the problem, where good layouts to small and medium problems are obtained and conclusions about the metaheuristic performance for big problems are established [5]. Gas outperforms other heuristic methods. GAs start by producing initial population and stop by reaching sub-optimization via genetic operators such as reproduction, crossover and mutation [7]. Each individual in a population acts once and in parallel to avoid falling onto a local optimal domain. The possibilities for reaching a global optimization are increased. Genetic algorithm is one the most frequently techniques inside the literature as method to solve combinatorial optimization problems and are considered to be an effective and robust approach. When solving the FLPS using Gas, each individual the population represents one facility layout.

* Corresponding author - Tel: +86-1366-820-2569;
E-mail: jinmaozhu @scu.edu.cn

The fitness of each facility layout is found, and as always, the best individuals get to pass the chromosomes on to the next generation. David M. Tate etc. [8] describes a GA approach to QAP that finds solutions competitive with those of the best previously-known heuristics, and argue that genetic algorithms provide a particularly robust method for QAP and its more complex problems. In addition, Islier etc. presents a model based on GA to facility layout using a load factor, a shape factor and a deviation factor to build the objective function [3]. This model has been widely used. For a Facility Layout Problem variation, that is the multi-channel manufacturing system design, Ozcelik and Islier propose a model based on GA to get better solutions for a problem found in the literature and found solution for a real life problem [9]. The technique chosen for this study was Genetic Algorithms with an improvement and construction approach.

This paper describes the adaptation of GA to the Facility Layout Problem. In Section 2 some generalizations about genetic algorithms are presented. The problem description and the model proposed are described in Section 3. The description of the proposed GA strategies, the chromosome design and the implementation are showed in Section 4. A genetic algorithm for the facility layout of a virgin office is proposed in Section 5. An experimental design to evaluate the algorithm efficiency is described in Section 6. The computational results and a comparison between actual and obtained layouts for each instance are presented in Section 7. Finally, conclusions are given in Section 8.

2 Genetic algorithms

Genetic Algorithms (GA) are random directed search techniques [10]. It was invented by John Holland, professor of the Michigan University, who published his article "Adaptation in Natural and Artificial Systems" in 1975. Genetic Algorithms (GAs) are the heuristic search and optimization techniques that imitate the natural selection and biological evolutionary process. They arose as an alternative form to the gradients methods, which could result little practical when searches in a n-dimensional space are made and multiple relative optimal are present [11]. They are inspired on the evolution and natural selection processes. GAs combine the notion of survival of the fittest, random and yet structured search, and parallel evaluation of nodes of search space [12]. A genetic algorithm consists of a string representation (chromosomes) of nodes in the search space, a fitness function to evaluate the search nodes, and a stochastic assignment to control the genetic operators. Thus, their operators, which are based on evolutionary methods, act on the individuals pertaining to each generation in order to find solutions that represent improvements in the objective function. The individuals (pertaining to the population) within an AG are feasible solutions of the

problem. The population is a subset of the space solution and its size depends on the treated problem. Each individual of the population is represented by a chromosome. This is a mixture of symbols known like genes. The basic representation methods of the individuals are binary and integer or real numbers vectors. The construction of a genetic algorithm for FLP can be separated into the following and yet related tasks:

- Coding of a facility layout.
- Determination of fitness functions.
- Designing of genetic operators.
- Implementation of GAs.

A challenging problem in applying the GA is the coding of solutions of the facility layout as chromosomes. Take care to ensure that the meaning of original problem is preserved under the coding scheme. Commonly used representations are non-binary chromosomes, which consist of non-binary genes and are often called enumerated representation. In spite of the difficulty in codifying the individuals, to design the search strategies in which the allocation of the specific parameters of the problem is fundamental and the management of the constraints of the problem through genetic algorithms, it is possible to obtain good distributions from a guided search on a small fraction of the solutions space [13].

3 Problem formulation

Theme problem consists in establishing the optimal layout for four offices of a real tourist service centre network, through a computational implementation using a metaheuristic technique, with the purpose of increasing the efficiency in the operation of the same ones. The metaheuristic selected was Genetic Algorithms. The offices considered in this study were chosen because they were representative of different branches types. The selection process was in charge of a team of people that is part of the commercial management of the tourist service centre. The first of the four offices of the financial organization considered in this study consists of two plants. That is the reason why in the development of this work, each one of these will be considered like an independent office. Hence, five offices will be handled, and these will be the instances of the problem. The measures of the office and the departments corresponding to each studied branch were obtained from the planes provided by the tourist service organization. From interviews made to the commercial and operative managers of the tourist service centre network, surveys were obtained in which the relative importance of proximity between the departments of each office was graded. This set of grades, that from now on we will call perception matrix, are analogous to the measurement of the flow. Given that, the greater the flow between a pair of departments, the greater will be importance of proximity between them, and then the qualification in the perception matrix will be proportional to the flow

between departments. The format used is a variation of the relation activities diagram proposed by Muther [14].

The model proposed by Islier does not applied to the problem treated in this project, since although it is wanted to minimize the work load; a unit costs matrix is not available. On the other hand, a fixed area for each department exists, since the dimensions of the facilities are established from the beginning of the layout design process, unlike the model by Islier where minimum and maximum wished areas of each department are established. Thus, the following model is proposed:

Given the problem's nature, the objective function is:

$$O.F. = \sum_i \sum_j [f(i, j) * d(Z(i), Z(j))], i, j \in [1, N].$$

Subject to:

- The departments are not overlapped to each other.
- The departments and office area is respected.
- The flank to which each department belongs is respected.
- A relative fixed position for some departments is conserved.

Where:

N is the number of departments.

$f(i, j)$ represents the flow between department i and department j .

$d(Z(i), Z(j))$ represents the distance between the rectangles' centroids, where department i and department j are located, according to the allocation defined by the chromosome.

This distance is measured using rectilinear metric, thus:

$$d_{ij} = |x_i - x_j| + |y_i - y_j|, i, j \in [1, N],$$

where:

d_{ij} is the distance between department i and j .

x_i is the horizontal coordinate of the centroid of department i .

y_i is the vertical coordinate of the centroid of department i .

x_j is the horizontal coordinate of the centroid of department j .

y_j is the vertical coordinate of the centroid of department j .

4 Genetic algorithms for the facility layout problem

The problem solution is represented in a chromosome that is divided in five parts. The first four ones represent the flanks of the office (West, North, East and South), the fifth one represents the departments which the client does

not have direct access and they are in the office back part. Each one of the first four parts of the chromosome contains the numbers of the departments in the order in which they are located. The ordering of departments within the plane is made in the sense of the movement of the clock's hands. The fifth part contains the numbers of the departments located in the back part of the office, in form of "columns", which go from left to right, and are filled from top to bottom. For greater clarity we present the following example:

The chromosome:

1	3	2	13	14	9	12	15	16	11	6	4	5	8	7	10	18	17
---	---	---	----	----	---	----	----	----	----	---	---	---	---	---	----	----	----

Represents the layout in figure 1.

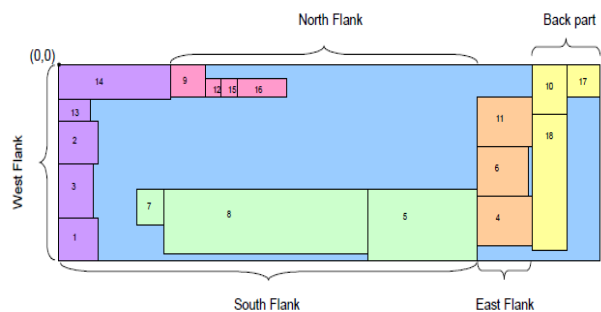


FIGURE 1 Layout corresponding to the representation of the previous solution

In order to generate the initial population, to each one of the parts of the previous chromosome, a random number of interchanges between departments is made, as long as there is no constraint forbidding the interchange. From this process a new chromosome is generated that along with others generated of the same way, conforms the initial population.

The fitness evaluation of each individual consists in the sum of the product from the flow between departments and the rectilinear distance between the centroids of these. The selection strategy used was elitist where the best individual in cost terms of a generation goes intact to the following one.

The used strategy of crossing is uniform. After making the crossing between two chromosomes, the feasibility of the generated sons is verified, so that no department is absent within the chromosome. If the chromosome needs to be repaired in some of its parts, the first one repeated is replaced by the first absentee, in the departments that belong to that part.

The mutation operator works of the following way: Two random numbers between 1 and the number of departments that belong to the part of the chromosome that is being muted are obtained, and the positions corresponding to these numbers in that chromosome's part are interchanged. The mutation is made by each one of the five chromosome parts, and departments are interchanged between them only if these have not been assigned like fixed within the configuration previously defined by the user. Thus, it is guaranteed that the chromosomes continue being feasible after mutation.

Once the population for each generation is obtained, the algorithm evaluates the Objective Function for each one of the chromosomes. Because of this, it is necessary to find the coordinates of each department according to the ordering represented for each one of the chromosomes. New coordinates are found for all the departments of the chromosome. The departments assigned as fixed, conserve their position within the vector and although their coordinates could vary according to the new allocation, they conserve a fix relative position within the layout. The origin for the measurements is in the left superior corner of the office's plane.

The following pseudo code describes the way how the new coordinates are found. *InX* is a variable that keeps the position in *X* coordinate and *InY* has the same function for the *Y* coordinate. *InXnow* is a variable that keeps the position in *X* coordinate for the back part of the office.

```

Algorithm to Find coordinates;
InY = Office height
InX = 0
For all departments ∈ West flank do
X coordinate = 0
Y coordinate = InY - Department height
InY = Y coordinate
Next
InY = greater departments height of the North flank
InX = greater departments wide of the west flank
For all departments ∈ North flank do
X coordinate = InX
Y coordinate = InY - Department height
InX = InX + Department wide
Next
a = greater departments wide of the East flank
b = Office wide - Back part wide
InX = b - a
InY = greater departments height of the North flank
For all departments ∈ East flank do
X coordinate = InX
Y coordinate = InY
InY = InY + Department height
Next
InY = Office height - greater departments height of the Costado Sur
For all departments ∈ Costado Sur do
InX = InX - Department wide
X coordinate = InX
Y coordinate = inY
Next
    
```

FIGURE 2 Pseudo code to find coordinates

5 Genetic algorithms for the facility layout of a start-up tourist service centre

In order to establish the facility layout of an office from zero, a metaheuristic model based on GA, similar to the previous one, is proposed. The construction, operators and output are equal to the explained ones in the previous algorithm, except in the generation of the initial population, which was generated randomly, considering the flanks to which each department belongs, according to the information entered previously by the user.

6 Experimental designs

With the objective to establish the best combination of entrance parameters for the algorithm, for each one of the considered instances, three parameters, each one with three levels will be varied: Mutation rate (α): 0.01, 0.05 and 0.1, best offspring generated fraction (β): 0.3, 0.6 and

0.8, and best fathers fraction (γ): 0.2, 0.45 and 0.7, with a size of population of 100 and 1000 iterations. Throughout 30 replications, the minimum, average and maximum values for the objective function obtained are shown in table 1.

TABLE 1 Minimum, average and maximum values for the objective function obtained throughout 30 replications, for all the parameters combination

combination	α	0.01	0.01	0.01	0.05	0.05	0.05	0.1	0.1	0.1	0.01	0.05	0.1
	β	0.3	0.6	0.8	0.3	0.6	0.8	0.3	0.6	0.8	0.6	0.8	0.8
	γ	0.2	0.45	0.7	0.2	0.45	0.7	0.2	0.45	0.7	0.7	0.45	0.2
Instant 1	Minimum	4413.050											
	Mean	4413.050											
	Maximum	4413.050											
Instant 2	Minimum	617.100											
	Mean	617.100											
	Maximum	617.100											
Instant 3	Minimum	4207.250											
	Mean	4207.250											
	Maximum	4207.250											
Instant 4	Minimum	3700.375											
	Mean	3700.375											
	Maximum	3700.375											
Instant 5	Minimum	2702.125											
	Mean	2702.125											
	Maximum	2702.125											

By the experimental design showed previously, it is possible to affirm that the algorithm is robust, since in spite of the variations in the number of iterations, both the mean and the minimum value obtained present great stability in contrast to the variations of the parameters. At the same time, the maximum value that is obtained for a great number of experiments can vary according to the number of iterations. Thus, it is advisable to consider that a mutation rate superior to 0.05, will throw very good solutions to the studied problem.

7 Results

The facility layouts given by the algorithm are approaches on the optimal physical distribution of the same one, so that they are aids for the layout designer and at no moment they try to be an exact representation in the architectonic sense of the floor of the plant. In the following table, the number of departments, the objective functions values obtained for the present distributions and the best distributions found by means of the algorithm for each office are compared.

TABLE 2 Comparison between the present distribution and the best one found by the algorithm for each office, in terms of its F.O.

Office	Number of departments	Present Layout O.F.	Proposed Layout O.F.	Improvement percentage ¹⁸
1	18	4,751,900	4,413,050	7.13%
2	8	622,400	617,100	0.85%
3	16	4,322,750	4,207,250	2.67%
4	16	3,765,275	3,700,375	1.72%
5	15	2,860,425	2,702,125	5.53%

From the previous table, it is clearly that in all the treated instances the objective function of the problem improved. From which it is deduced that the algorithm is able to design better distributions in spite of the limitations of space and number of feasible solutions that a small problem in terms of number of departments can represent.

8 Conclusions

As Tompkins states the services sector is often managed in a non-scientific way in spite of its size. The approach taken in this paper is the application of Operational Research concepts aimed to provide a more scientific solution to this particular problem [15]. The decision on the design of a more adapted layout is a multicriterial optimization problem. Because of this, it is recommendable to resort to experience, intuition and financial analysis techniques, according to the importance that the organization grants to each one.

A high generation's number is not an infallible method to always obtain the best well-known solution. Thus, in the search of good solutions for the studied problem, it is recommended to use a great number of generations, along with a mutation rate superior to 0.05. The algorithm developed is highly flexible, since starting from a present layout, fixed objects, prohibited zones, and contour office limitations can be considered, assuming to these two last ones as a fixed department within the plant. The distribution designs obtained by means of the genetic algorithm proposed satisfy in a better way the proximity necessity between the departments that must be next. According to the algorithm performance observed in each instance, it is possible to state that the algorithm is robust, since in spite of the variations in the number of iterations and entrance parameters, the average presents a remarkable stability, and in all cases the minimum known for each instance of the problem was obtained.

The scope of this paper should be extended to construct robust models that solve the facility layout problem in a tourist service centre considering other restrictions that were not considered in this work, such as Offices that have several floors, nonrectangular office plant, variable plant size, unknown plant size, prohibited zones within the hall destined for the clients, minimum and maximum areas for the departments and three-dimensional design. As alternative solution techniques, other metaheuristics like Simulated Annealing, Tabu Search, etc., and Hybrids methods integrated with Local Search strategies could be considered. The last one helps




to increase exploration and to avoid solutions that correspond to local optimal. Additional to this, it turns out interesting to solve the problem with multiobjective approach, generating a family of solutions in the efficient border. Finally, further research could be done to extend the application of these methods to any other kind of service industry.

Acknowledgments

This research was supported by the Major International Joint Research Program of the National Natural Science Foundation of China (Grant No. 71020107027), the National Natural Science Foundation of China (Grant No. 71001075), Humanities and Social Sciences project of The Ministry of education of China (Grant no.12YJC630023), Social sciences program of Chengdu (Grant no. SKNHR13-07).

References

- [1] Tompkins J A, White J A 1984 *Facilities Planning* John Wiley&Sons New York
- [2] Kusiak A, Heragu S S 1987 The facility layout problem *European Journal of Operational Research* **9** 229-251
- [3] Islier A A 1998 A genetic algorithm approach for multiple criteria facility layout design *International Journal of Production Research* **36**(6) 1549-1569
- [4] Chiang W C, Kouvelis P 1996 An improved Tabu search heuristic for solving facility layout design problems *International Journal of Production Research* **34**(9) 2565-2585
- [5] Mavridous T D, Pardalos P M 1997 Simulated annealing and genetic algorithms for the facility layout problem: a survey *Computational Optimization and Applications* **7**(1) 111-126
- [6] AI H L 2002 On solving facility layout problems using genetic algorithms *International Journal of Production Research* **3**(11) 573-2582
- [7] Wang K 2003 *Intelligent condition monitoring and diagnosis systems – A computational intelligence approach*. Frontiers in Artificial Intelligence and Applications, IOS Press, The Netherlands
- [8] Tate D M, Smith A E 1995 A Genetic approach to the quadratic assignment problem *Computers and Operations Research* **22** 73-84
- [9] Ozcelik F, Islier A A 2003 Novel approach to multi-channel manufacturing system design *International Journal of Production Research* **41**(12) 2711-2726
- [10] Gen M, Chang R 1997 *Genetic Algorithms and Engineering Design Wiley-Interscience*
- [11] Tam K Y 1992 Genetic algorithms, function optimization, and facility layout design *European Journal of Operational Research* **63** 322-346
- [12] Chan K C, Tansri H 1994 A study of genetic crossover operations on the facilities layout problem *Journal of Computers and Industrial Engineering* **26** 537-550
- [13] Michalewicz Z 1996 *Genetic algorithms + data STRUCTURES = evolution programs* New York: Springer
- [14] Muther R 1973 *Systematic Layout Planning* Boston: Cahnerns Books
- [15] Tompkins J A 2010 *Facility planning* John Wiley&Sons

Authors	
	<p>Wei Shang, born in January, 07, 1977, Chengdu</p> <p>University studies: Wei Shang received his Ph.D. in the College of Business Administration of Sichuan University in 2010.</p> <p>Scientific interest: His major research interests are platform marketing and product harm crisis.</p> <p>Publications: He has published more than 10 journal and conference articles in referred journals and conferences.</p> <p>Experience: He has often worked as a business consultant in more than 10 enterprises.</p>
	<p>Maozhu Jin, born in March, 14, 1978, Chengdu</p> <p>Current position, grades: Maozhu Jin is an instructor of Business School, the tutor of MBA operations management and innovation and entrepreneurship management in Sichuan University. He has been engaged in the teaching of core curriculums such as operations management and management consulting.</p> <p>Scientific interest: His current research interests include the areas of operations management, organizational process reengineering, strategic management, service operations management, platform-based mass customization and risk management.</p> <p>Publications: He has published two books and over ten research papers in referred journals of high quality both at home and abroad, and ten of them are retrieved by SCI and EI.</p> <p>Experience: As a main researcher, he has participated in and completed three projects supported by National Natural Science Foundation of China.</p>
	<p>Jie Wang, born in April, 14, 1991, Chengdu</p> <p>University studies: Jie Wang is now a graduate for master degree at business school of Sichuan university.</p> <p>Scientific interest: Her major research interests are tourism management and supply chain management.</p> <p>Experience: As a main researcher, she has participated in many projects supported by various funds of China.</p>

Method of traffic zone division based on spectral graph theory

Wu Shimei^{*}, Pei Yulong, Cheng Guozhu

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Received 1 March 2014, www.tsi.lv

Abstract

Aiming at the macro planning of traffic district division, the spectral graph theory, graph, spectral and matrix (i.e. Laplace matrix) were introduced, and the method and procedures of the traffic zone division based on spectral graph theory were put forward. Euclidean distance between communities was calculated according to four indexes, i.e., economic indicators, land development intensity, residential population quantity and spatial distance. The balance degree of indexes between communities was regarded as criterion to evaluate community division precision. Taking Dongguan, a city in the south China, as an example, above mentioned method was applied, conducting the traffic zone division, with rail transit network planning flow prediction.

Keywords: Traffic zone division, Spectral graph theory, Euclidean distance, Dongguan

1 Introduction

Traffic zone division is the basis and a key step of predicting the passenger flow. Traffic zones can be divided into macro planning, middle management and micro control [1]. Generally, the macro planning means the planners divide the areas in accordance with the administrative boundaries or roads, railways, mountains, rivers and other natural lines[2], with no comprehensive consideration of the usage of land, population, traffic and other factors within the community, and the result is comparatively subjective. This directly influences the traffic forecast precision. In order to improve the arbitrariness and limitations of traffic zone division at macro planning level, this paper introduces the spectral graph theory and proposes a new method of traffic zone division.

A comprehensive understanding of the traffic flow among the traffic resources is needed when processing the transportation planning management. However, because of the large amount of resources, it is impossible to research the resources individually. Therefore, integrating several traffic zones into one is required. Whether the divisions are proper is directly related to the research, analysis, prediction, labour force and precision. The smaller the traffic zone is divided, the more accurate it is. However, it takes much more labour forces with higher costs when the traffic zones are too small. On the basis of ensuring the accuracy of the research, taking advantage of the existing materials (like income, occupation, etc.), or the previous data given by the cities, which have been investigated on residents travel, and reducing the numbers of the traffic zones being researched, so as to decrease the workload, are difficult. In current China, the size of the traffic zone is determined by the land usage, population distribution, administrative

division, geography, and the layout of roads, etc. As for the radius of the traffic zones, no uniform and related standard or theory is at hand, but with some existing data.

2 Literature Review

The method of traffic zone divisions in America is in accordance with the area and population of the zones determined by the size of the scale of the urban area. Details are as follows: when the population is less than 715,000 within the zone, the average area is 1,138 km² with the average population of 872; when the population is less than 1,510,000, the average area is 2177km² with the average of population of 954; then the population is less than 10,000,000, the average area is 5155 km² with the average population of 2828; when the population is more than 10,000,000, the average area is 7183 km² with the average population of 7339. The method of traffic zone divisions in European countries is on the basis of population, for example, the regulation in Russia is as follows: with population of 100,000 ~ 250,000, the number of the zones should be 5~10; with the population of 250,000~500,000, the number should be 8~20; with the population of 500,000~1,000,000, the number should be 15~20; with the population of 1,000,000~2,000,000, the number should be more than 50. Clustering analysis is applied in Chinese related documents when research on the traffic zone divisions, which include: automatic classification method based on the spatial analysis was proposed by Li Xiaodan and Yang Xiaoguang [3]. The model of the clustering analysis algorithm was given in the document, planning the procession and arithmetic of the divisions. The method of divisions based on the fuzzy analysis was proposed by Zhao Jinhuan, Li Wenquan, etc. [4]. In accordance with the fuzzy clustering of the equivalence matrix in clustering analysis, by making use

^{*} Corresponding author - E-mail: 4770260@qq.com

of the data of ages, occupations, incomes and ratio of car ownership, programming the MATLAB, and using statistical magnitude F to determine the best fuzzy clustering number, the divisions of the traffic zones is realized. Dynamic division method based on the GPS data of the taxis was proposed by Lv Yuqiang [5]. A revised method based on the clustering analysis was proposed by Yang Bo and Liu Haizhou [6]. The method based on the clustering analysis was proposed by Liu Yifei [7].

The method of radius calculation of the traffic zones based on the proportion of travel inside the zones was proposed by Ma Chaoqun [8]. On the basis of the three different combinations of the radius and trip distance, the calculation model of the proportion of travel was elicited in the document, and the quantitative relation of the three was processed. The reasonable range of the radius of the traffic zones were produced by controlling the proportion of travel inside the zones. Double divided method was proposed by Dai Binggui and Hu Ji, which was based on the weighted modification of the fuzzy clustering analysis by adopting the level analysis, and established the method of weighted fuzzy clustering analysis to research the divisions of the traffic zones [9].

3 Spectral graph theory

Spectral graph theory is to study the collection of eigenvalue of a Laplace matrix or adjacency matrix in a graph. It is an effective approach to describe the diagram’s structural characteristic with its eigen value and characteristic vector as well as its algebra representation. By analysing the spatial features of graph, spectra graph theory establishes the topology map (especially the graph’s various invariants) and connection between the eigen values; By making use of algebraic theory, geometric theory and probability method, the nature of topology of a graph is revealed. Concepts related with Spectral graph theory that are discussed in this paper include graph, spectrum and Laplace matrix. [10].

3.1 GRAPH AND SPECTRUM

3.1.1 Graph

Graph G is composed of two sets, V and E , and $G=(V,E)$. V is a set of n nonempty vertexes, mathematically, it is expressed as $V=V(G)=\{v_1, v_2, \dots, v_n\}$; E is a collection of graph edges, mathematically expressed as $E=E(G)=\{(v_0, v_1), (v_0, v_2), \dots, (v_i, v_j), \dots, (v_n, v_{n-1})\}$.

3.1.2 Spectrum

λ is known as the n values of n -order matrix A . If there is any non - zero n -order vector which realizes

$AX = \lambda X$, then X is called an eigen vector of A corresponding with eigen value.

Known that, λ is the roots of polynomials $\chi_A(\lambda) = \det(\lambda I_n - A)$, in which I_n is an n -order unit matrix, and $\chi_A(\lambda)$ is known as characteristic polynomials.

According to Gauss's theorem, the characteristic equation $\chi_A(\lambda) = 0$ has n roots: $\lambda_1, \lambda_2, \dots, \lambda_n$. Since $\lambda_1, \lambda_2, \dots, \lambda_n$ are not necessarily distinct, it is called the spectrum of matrix A , renamed as $spec A$:

$$spec A = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_s \\ m_1 & m_2 & \dots & m_s \end{pmatrix}.$$

$\lambda_1, \lambda_2, \dots, \lambda_n$ are distinct, and m_i are the algebraic multiplicity of λ_i .

3.2 LAPLACE MATRIX

Laplace matrix is matrix L , which is a $n \times n$ dimensional symmetric matrix whose element L_{ii} on the diagonal line is the degree of node i , while the elements L_{ij} on other off-diagonal means the connection of i and j . Laplace matrix usually has three kinds of representations, respectively, $D-A$, $A-D$ and $D+A$, and the calculation formula are as follows:

$$L_{D-A} = [l_{i,j}] = \begin{cases} d_{ij} & i \neq j \\ -\sum_{k=1}^n d_{i,k} & i = j \end{cases},$$

$$L_{A-D} = [l_{i,j}] = \begin{cases} -d_{ij} & i \neq j \\ \sum_{k=1}^n d_{i,k} & i = j \end{cases},$$

$$L_{D+A} = [l_{i,j}] = \begin{cases} d_{ij} & i \neq j \\ \sum_{k=1}^n d_{i,k} & i = j \end{cases}$$

d_{ij} is the Euclidean distance between two feature points of G .

If G is connected, then the second smallest eigen value λ_2 of Laplace matrix is positive, and its corresponding eigen vector (called Fiedler vector) contains important information of the graph, i.e. the value of elements or (including positive and negative ones) reflects the relationship between the corresponding vertices.

4 The method of traffic zone division

4.1 THE PROCESS OF DIVISION

Making use of the spectral graph theory, the process of division is shown as follows:

- 1) Primarily divided the study area into subareas according to the administrative division and number them. The size of planed area should be considered and comparatively small administrative division should be applied (such as small village, town or county administrative division);
- 2) Calculate and normalize the Euclidean distance in appliance with the index value and construct the Laplace matrix.
- 3) Calculate the second eigen value λ_2 of Laplace matrix and the corresponding eigen vector (Fiedler vector)
- 4) Integrate the original traffic zones into two, satisfying the value of each element of eigen vector.
- 5) New residential parks containing most original traffic zones are selected to determine whether the number of original traffic zones is greater than the fixed M;
 1. If it is larger than M, establish the Laplace matrix for the new community, and return to step 3) for further division of the community;
 2. If it is smaller than M, stop the division.

The value of M is determined by the preliminary division of the traffic zone's administrative and planning level, as well as the accuracy. For example, urban traffic planning if applied, and the planning is required to be as detailed as traffic facilities of inner-countries and intra-countries., then the preliminary divided traffic zone should be limited within the county, in accordance with the administrative divisions below the county level (such as town). If there are 20 towns in average, then the minimum M can be set as 20 or below. Then, in the (4th) step, according to the value of elements of eigen vector, the original traffic zone is divided into two parts in the following way: choose the critical value S of eigen vector $F = (f_1, f_2, \dots, f_n)$ according to certain rules, and the original traffic zone corresponding to $f_i \geq S$ is regarded as one traffic zone, and the rest is set as another traffic zone. The ways to determine the critical value of S are as follows:

1. Bisection method. To order the elements in eigen vector (Fiedler vector) from small to large (or from large to small), and the value of the element in the middle is set as the critical value, i.e., the new traffic zones contain the same number of the original zones.
2. The average method. It can be proved that the sum of all elements of eigen vectors is zero, so is their mean, i.e. $S = 0$.
3. The method of maximum distance Order the elements in eigen vector (Fiedler vector) from small to large (or from large to small), and calculate the distance

between two adjacent elements, then set the maximum distance as S .

4.2 CALCULATION OF EUCLIDEAN DISTANCE

Euclidean distance is an arithmetic that calculates the distance between two points in space. In this paper, the Laplace matrix is constructed by making use of the Euclidean distance calculated by various index. Firstly, make sure each traffic zone shares the same index. Knowing from above analysis, the factors affecting the division of the community are economic indicators, land development intensity, population of the community, area of community and the distance between the communities. Taking the four as examples, suppose that the economic index of the community i is A_i , the land development intensity is B_i , the population is set as C_i , area is set as D_i , and the spatial distance is set as L_{ij} , then the Euclidean distance between the two different communities can be expressed as [11]:

$$d_{ij} = \sqrt{(A_i - A_j)^2 + (B_i - B_j)^2 + (C_i - C_j)^2 + (D_i - D_j)^2 + L_{ij}^2}$$

d_{ij} is the Euclidean distance between i and j , and the Euclidean distance between each two communities can be obtained, then Laplace matrix based on Euclidean distance can be constituted as shown above.

Economic index A_i , population C_i and area D_i can be obtained by the mean of the number of communities divided by all communities; the index of land development intensity B_i is obtained by floor area ratio (FAR) of this zone divided by that of the whole studied area; spatial distance index L_{ij} is the value of the distance between geometric centres of two residential parks divided by the radius of a circle with the same area in the researched area [12].

4.3 CALCULATION PROCEDURE

Generally, the iterative calculation of matrix adopts MATLAB for programming for now. The MATLAB software provides a comprehensive solution for many scientific fields that need numerical calculations by integrating the numerical analysis, matrix calculation, data visualization, modelling and simulation of nonlinear dynamic systems and many other powerful functions into an easy-to-use Windows environment, [13] This paper, using MATLAB to program, calculate the Euclidean distance, eigenvalue, eigen vector, etc., between every two communities for each step.

5 Evaluation of division accuracy

The feasibility and validity of these traffic zones divided by using above methods should be evaluated according to

certain indexes, because the division has an important influence over the subsequent traffic forecast and traffic distribution, the appropriate division of traffic zone is the key point of the research. The following criteria can be adopted for evaluation:

(1) The population of each traffic zone being finally divided should keep a balance. In order to ensure the precise value of traffic flow between traffic zones, the population of each traffic zone should not differ too much. Assuming the suitable population is P , the population of each community is $P_i (i=1,2,3..n)$, the allowed error limit is ΔP , and then the population should satisfy the following relations:

$$\begin{cases} |P - P_i| \leq \Delta P (i=1,2,3..n) \\ |P_i - P_j| \leq \Delta P (i \neq j) \end{cases}$$

(2) The uniformity of area of traffic zones: As a key factor for traffic zone division, the size of the areas not only influences the internal travel probability, but also affects the flows among traffic sources. Assume that suitable area is D , the subarea is $D_i (i=1,2,3..n)$, the allowed error limit is ΔD , then the area of residential park should satisfy the following relations (if a residential park contains hills, waters, then that area shall be deducted):

$$\begin{cases} |D - D_i| \leq \Delta D (i=1,2,3..n) \\ |D_i - D_j| \leq \Delta D (i \neq j) \end{cases}$$

(3) The economic level of all traffic zones should reach a certain standard equally. The economic level is an important factor to affect traffic generation and attraction. Developing area usually embraces little transportation, while developed area embraces large traffic flow, with huge travel and attraction. Therefore, the economic level of all traffic zones should be kept at an appropriate level.

6 Case studies

Taking Dongguan rail transit network planning and passenger flow forecast as an example, analysing the divisions of the traffic zone by making use of the above-mentioned methods. Dongguan, borders among Guangzhou, Huizhou and Shenzhen, covers a total area of 2465 km², with 35 town-level administrative units and 591 village-level administrative units, which are in jurisdiction. Dongguan, due to historical reasons, has no county-level administrative structure as the usual administrative regions would have, but has 35 town-level administrative units directly under jurisdiction.

6.1 PRELIMINARY DIVISION OF THE TRAFFIC ZONE

Dongguan holds 35 town-level administrative units and 591 preliminary divided traffic zones with 16.8 town-level administrative unit in average. In order to appropriately improve the prediction accuracy, M is set as 10.

6.2 PROCESS AND RESULTS OF TRAFFIC ZONE DIVISION

In this paper, the most economically developed south-western area is taken as an example for traffic zone division according to the above method. The south-western area embraces 4 town-level administrative units, including Houjie, Shatian, Humen and Changan, and 84 village-level administrative units. According to the 2011 Dongguan statistical yearbook and regulatory detailed planning of the four towns, the data of population, GDP and land development intensity of the 84 village-level administrative units can be easily obtained. See Table 1.

TABLE 1 Data summary of some village-level administrative

Village name	The town it is located in	GDP (billion CNY)	FAR	Population	Area (km ²)
Chenwu	Houjie	0.96	1.7	11503	2.0
Mintian	Shatian	0.48	0.8	14028	5.9
Zexu	Humen	1.05	1.4	32764	2.5
Wusha	Changan	2.58	1.9	112356	10

In 2011, the average population in the 84 village-level units is 21,022, with the average GDP of 780,000,000 CNY, the average FAR of 1.1, and the average area of 6.03 km², The radius of the circle with the same area is 12.7km.

The steps of division of traffic zones are as follows:

- 1) Normalizing the data, and calculate the index of the population, economic development, land development -intensity and the area, (calculate the spatial distance respectively) of the village-level administrative units seeing table 2.

TABLE 2 Village administrative unit index summary

Village name	The town it is located in	A_i	B_i	C_i	D_i
Chenwu	Houjie	1.23	1.55	0.55	0.33
Mintian	Shatian	0.62	0.73	0.67	0.98
Zexu	Humen	1.35	1.42	1.56	0.41
Wusha	Changan	3.30	1.73	5.34	1.66

- 2) Calculate the Euclidean distance among villages, and create the Laplace matrix. See Table 3.

TABLE 3 Euclidean distances for some village-level administrative

Village name	Chenwu	Mintian	Zexu	Wusha
Chenwu	0	1.3846	1.2788	5.5594
Mintian	1.3846	0	1.9467	5.7084
Zexu	1.2788	1.9467	0	4.5508
Wusha	5.5594	5.7084	4.5508	0

- 3) According to the calculation, $\lambda_2 = 1.6$, the corresponding eigen vector (containing 84 elements) can be calculated, and its critical value = 0 derived from the average method;
- 4) At first, combine the divided zone into two traffic zones, with 38 and 46 village-level administrative units respectively.

5) Then divide the latter one into two subzones, with 25 and 21 village-level administrative units individually.

6) Finally, divide the area into 10 subzones. (See Figure1).



FIGURE 1 Traffic zone division process of Dongguan south-western area

With the similar steps, Dongguan is divided into 66 traffic zones as a whole. As a result, some traffic zones break the town-level administrative region, (for example, the No. 05 traffic zone is composed by some village-level administrative units of Houjie Town and Humen town), and they are connected with Dongguan town-level administrative units for continuous development, forming a dense urban cluster pattern, with closely related economy and traffic with adjacent towns.

6.3 EVALUATION OF DIVISION ACCURACY

In this paper, the average population and area of the 66 traffic zones are set as standard, (not including the large-scale mountain and water), and the allowing error is controlled within 15%. At the same time, Dongguan is a prefecture-level city, with medium-sized area, but its quantity of population and economic development level are above the medium city scale, so it is reasonable that the traffic zones are more than 50.

According to the requirements of planning administrative departments, the largest difference between traffic zones in population, GDP and area should be controlled within 30%. As is calculated, the largest difference of population is 13%, between NO.05 and No .07; largest GDP difference is 22%, between No.17 and No, 29; largest area difference is 28%, between No.33 and No.64. , Thus, the precision of division is totally in accordance with the planning requirements.

7 Conclusions

Traffic zone division is the priority and the most basic step for all kinds of transport planning and passenger flow prediction, and the results are directly related to the precision and difficulty of the passenger flow prediction. Through the introduction of spectral graph theory, a new method for traffic zone division is proposed. The

divisions of the traffic zones are realized by adopting the key factors that affect the divisions, normalizing the index, calculating the Euclidean distance and creating the Laplace matrix. Dongguan, as an example, is divided into several traffic zones according to this method, and the division results conform to the planning administrative departments' requirements in precision.

References

- [1] Li Xiaodan 2009 *Theory and method of city road network traffic zone division* Doctoral dissertation of Tongji University
- [2] Lu Huapu 2006 *Theory and method of traffic planning* Beijing: Tsinghua University press
- [3] Li Xiao-dan, Yang Xiao-guang, Chen Hua-jie 2009 Study on traffic zone division based on spatial clustering analysis *Computer Engineering and Applications* 45(5) 19-22
- [4] Zhao Jin-huan, Li Wen-quan 2009 Improvement of the Traffic District Partition in Resident Trip Investigation *Journal of Transportation Engineering and Information* 7(2) 110-115
- [5] Lu Yu-qiang, Qin Yong, Jia Li-min etc. 2010 Dynamic Traffic Zone Partition Based on Cluster Analysis of Taxi GPS Data *Logistics technology* 2010(216) 86-89
- [6] Yang Bo, Liu Hai-zhou 2007 Improvement of the Method about the Partition of Traffic Zone *Traffic and Transportation* 7
- [7] Liu Yi-fei 2011 Traffic District Division Method and Application Based on Fuzzy Cluster Analysis *Logistics Sci-Tech* 2011(9) 25-28
- [8] Ma Chao-qun, Wang Rui, Wang Yu-ping etc. Calculating Method of Traffic Zone Radius in City Based on Inner Trip Proportion *Journal of Traffic and Transportation Engineering* 7(1) 68-72
- [9] Dai Bing-kui, Hu Ji 2010 Double-level Division Method of Traffic Zone *Railway Transport and Economy* 32(11) 86-89
- [10] Zhuang Zhenhua 2010 *Data classification of gene expression spectrum based on Laplace spectrum* Anhui University master's paper
- [11] Chung F R K 1997 *Spectral Graph Theory* American Mathematical Society Providence Rhode Island
- [12] Luo Fengtao 2012 Application of Euclidean distance matrix analysis (EDMA) in in medical morphology research *Anatomy research* 34(1) 50-54
- [13] Chen Huaichen, Wu Dazheng, Gao Xiquan 2003 *MATLAP and its application in electronic information courses* Beijing: Publishing house of electronics industry

Authors	
	<p>Wu Shimei, born in July 13, 1978, China</p> <p>Current position, grades: School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, P. R. China. University studies: Ph.D. candidate of transportation science and engineering Scientific interest: urban planning, transportation planning, traffic management, traffic demand forecasting, etc. Publications: 10 publications Experience: He has often worked as a business consultant in more than 10 enterprises.</p>
	<p>Pei Yulong, born in October 20, 1961, China</p> <p>Current position, grades: Traffic College, Northeast Forestry University, professor University studies: Ph.D. in Transportation Planning and Management(2002, Southeast University) Scientific interest: transportation planning, transportation safety, traffic control etc. Publications: 300 scientific papers and 18 books</p>
	<p>Cheng Guozhu, born in August 16, 1977, China</p> <p>Current position, grades: Ph.D. in Road and Railway Engineering(2007, Harbin Institute of Technology) University studies: School of Transportation Science and Engineering, Harbin Institute of Technology, associate professor Scientific interest: transportation safety Publications: 70 scientific papers and 6 books</p>

A dynamic optimization model on decision-makers and decision-layers structure (DODDS) in C2-organization

Yanghe Feng^{*}, Baoxin Xiu, Zhong Liu

College of Information System and Management, National University of Defence Technology (NUDT),

Changsha, 410073, China

Received 1 March 2014, www.tsi.lv

Abstract

The highly-complexity, environmental uncertainty, and structure changes bring more requirements for the agility and resilience of its core command and control C2-organizations. In order to better understand such organization's dynamic and emergence behaviours as a system of systems, we establish time-domain based metric model to evaluate C2 organizational decision-making capability. We develop an optimization model of organizational structure. The model is based on decision-makers and decision layer dynamics. The model aims at helping gain an optimal organizational structure with higher operational flexibility, low cost and high performance.

Keywords: Operational System of Systems (SoS), Command and Control, Adaptive Optimization, Simulated Annealing

1 Introduction

The characteristics of system antagonism in war are more prominent due to its high complexity and dynamic uncertainty in information warfare. It is also due to contemporary war styles such as network-centric warfare and modern organization models like "Power to the Edge" [1, 2]. The models represent new principles of operations, enabling advanced organizational resources and potential features. The models also help adapt the environmental uncertainty and complexity by changing and evolving the system structure and its behaviours. This way, the organization can accomplish more complex tasks with a low cost and high performance even in worse and complicated situations. In military management science, such models can be equipped with the "brain" and "hub" functions to induce an information-centric warfare. The functions are established by implementing Command and Control (C2) ingredients to an organization. The C2-organizations take in charge of many critical tasks such as Trend Observation (TO), Information Processing (IP), Decision Making (DM) and Command Operations (CO), etc. For a C2-organization, the ability to dynamically optimize and quickly change its structure is critical to obtain the advantages of antagonism in today's information-centric war. Therefore, the dynamic optimization becomes one of the key issues in the System of Systems (SoS)-based C2 organization. Organization characteristics or behaviours like complexity, uncertainty and dynamics [1] should be carefully considered in a SoS Engineering (SoSE) perspective.

Based on the theories of organization design, most successful organizations become more flexible during its structure evolution. An efficient organization should dynamically adjust its structure according to environment changes [3, 4] in order to improve and increase its operational performance. However, such adjustment may increase its technical complexity [5, 6]. Alternatively, people may adopt a more authoritarian model to protect outside changes. Such model's organization is framed in terms of a hierarchy network. However, such defensive model, on the other hand, weakens its structure [7] and become a close system. In many models, the organizational structure, centralized/decentralized decision-making mechanism, and performance outputs were intensively investigated under various conditions of different complexities and environmental uncertainties. The various system relationships between environmental changes and organizational structure were studied by corresponding computer simulations [8]. The simulation models were used to investigate various influences of environmental uncertainty, decision-making module and operational structures on the system performance. Unfortunately, such impacts have not been rigorously addressed in both temporal and spatial dimensions. Such transient influences (dynamic impacts) due to the system environmental changes in time-space domain should be considered in a contemporary model of C2 systems.

To address this issue requires a strong background and knowledge in multi-disciplines such as computational and mathematical organization theories [9-12], system of systems engineering [13-15], information and organizational management [16, 17] etc. For example, through a series of US-based scenarios and the Adaptive

^{*} Corresponding author- fengyanghe@nudt.edu.cn

Architectures for Command and Control (A2C2) experiments, investigators developed many outstanding frameworks to establish various relationships in a C2-organization, hence to quickly build forces organizations in battlefield space. A typical framework commonly contains a multi-stage approach. The framework employs a series of algorithms developed to rapidly-construct military units in a battlefield. By using an adaptive optimization process, the organizational mission planning is divided into multiple stages. The degree of matching participated organizational structures is measured at all stages. The optimal adjustment of structural strategy is proposed throughout the structure changes, cost, and performance. For example, Liu et al. [3] grouped the similar resources and tasks together to facilitate the planning and tasks implementation. The classification is based on the modern group technology such as Nested Genetic Algorithm (NGA). In this way, they can simultaneously accomplish their missions on tasks-to-platforms and the allocations by platforms-to-decision-makers. Krackhardt and Carley proposed an improved organizational model, called Precedence, Commitment of Resources, Assignment, Network and Skill (PCANS) [13]. The model focuses on the description of information and their exchanges in the organization. They studied how to design an information exchange structure to satisfy the C2 requirements.

Unfortunately, most measures of C2 organizational-decision-making-capacity is only based on the entire mission period. In other words, in their models a large time scale has been used throughout the model simulations. Obviously, the models often fail to observe the mission performance during a transience or short period of time. That is somehow explain why these models in sometimes cannot well maintain organizational capability [14-16]. People start to consider how to divide a large mission time is divided into short ones; thus, the time-domain should be considered at small scales precisely in order to describe system dynamics. This consideration arise an important issue, i.e. is how to create a dynamic and optimal model that can instantaneously measure and rapidly adjust C2 organizational structure from time to time.

Motivated by the foregoing thinking, we propose a transient optimization model and its corresponding algorithms to measure and simulate dynamic C2-organizational structure. The model is built based on Dynamic Optimization of Decision-makers Decision-layer Structure (DODDS) in C2-organization to be depicted in the following. The model utilized an optimal time-domain division on the transient execution status of the operational tasks.

The rest of the paper is organized as follows: Section 2 introduces the C2-organizational system terminology and mechanism to measure its decision-making capacity. Section 3 presents the proposed dynamic structure optimization model. A brief conclusion is provided in Section 4.

2 Measure of decision-making capacity for C2 organization

The A2C2 experiments and adaptive organization designs [19, 23] commonly include decision-makers work load, communication and relationships. The C2 organizational elements can simply be classified into Platforms (P), Decision-makers (DM) and Tasks (T), respectively. C2 organization's intelligent layers can be divided into Decision-Making Layer (DM-L) and Scheduling Layer (S-L). DM-L considers two relationships: the hierarchical relationships between DMs (R_{DM-DM}) and the control relationships between DM and platform (R_{DM-P}). S-L is composed by the allocation relationships between platforms and tasks (R_{P-T}) and the sequence relationships for task allocation (R_{T-T}).

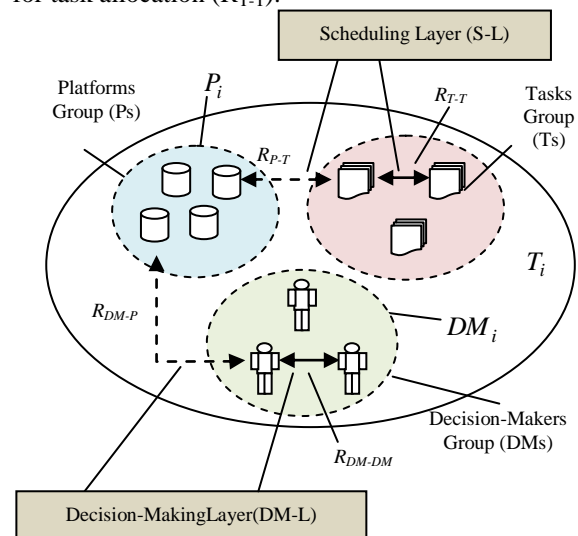


FIGURE 1 Illustration of C2-organization composition (dash arrow line stands for control relationships and solid arrow line refers to the command relationship)

The series of decision-makers in DM group can be denoted as $\{DM_i, i=1,2,3,\dots,M\}$. The subscript i for the i -th DM. M is the total C2 decision-makers. The platform set can be expressed as $\{P_j, j=1,2,3,\dots,N\}$, where is P_j the j -th platform. N is the total number of C2 platforms. Similarly, a set of tasks in C2 can be referred as $\{T_k, k=1,2,3,\dots,L\}$, where T_k is the k -th task. The total number of tasks under consideration is L . They can be expressed in terms of vectors as

$$DM = \begin{bmatrix} DM_1 \\ DM_2 \\ \vdots \\ DM_M \end{bmatrix}, P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{bmatrix}, \text{ and } T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_L \end{bmatrix} \quad (1)$$

These scale parameters are used to describe the organizational structure characteristics. Each element in the vector represents an organizational entity. The structure with different element setting may lead to different organizational effectiveness. The relationships among decision-makers, platforms and tasks can be

catalogued into five relationships. They are: (1) control relationship R_{DM-P} between decision-maker and platform, (2) allocation relationship R_{P-T} between a platform and a task, (3) hierarchical relationship R_{DM-DM} between two decision-makers, (4) the execution relationship R_{DM-T}^E between policy-maker and task, and (5) the command relationship R_{DM-T}^C among tasks. The fourth and fifth ones are two induced relationships through the DM-P and P-T relationships. The structure can be mathematically expressed as a group set $G_{OR} = (R_{P-T}, R_{DM-P}, R_{DM-DM}, R_{DM-T}^E, R_{DM-T}^C)$. These variables become physical quantities to measure the structure and its relations. The expressions of these variables are introduced in the following.

Platforms (such as hardware, software, devices, tools, facilities etc.) can usually be allocated to a special task. Such allocation relationships define the connections between platforms and tasks. It basically builds a network system. For example, one single platform can be used in multiple tasks, while a task often requires multiple platforms. According to [17-19], these C2 allocation relationships can be mathematically expressed as follows:

$$R_{DM-P}(i, j) = \begin{cases} 1 & \text{if } P_j \text{ is allocated to } DM_i \\ 0 & \text{else} \end{cases}, \quad (2)$$

where DM_i is the i -th decision-maker and j is the j -th platform. The control relationship matrix can be given as:

$$R_{DM-P} = [R_{DM-P}(i, j)] = \begin{bmatrix} R_{DM-P}(1,1) & R_{DM-P}(1,2) & R_{DM-P}(1,3) & \cdots & R_{DM-P}(1,N) \\ R_{DM-P}(2,1) & R_{DM-P}(2,2) & R_{DM-P}(2,3) & \cdots & R_{DM-P}(2,N) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{DM-P}(M,1) & R_{DM-P}(M,2) & R_{DM-P}(M,3) & \cdots & R_{DM-P}(M,N) \end{bmatrix}_{M \times N} \quad (3)$$

For example, if there are 3 decision-makers and 4 platforms in the C2-organization. $M=3$ and $N=4$. DM_1 has P_2 and P_4 . DM_2 has P_1 and P_2 . DM_3 has P_1, P_3 , and P_4 .

Then R_{DM-P} matrix is $R_{DM-P} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}_{3 \times 4}$.

Similarly, the allocation relationships can be mathematically expressed as follows:

$$R_{P-T}(j, k) = \begin{cases} 1 & \text{if } P_j \text{ is allocated to } T_k \\ 0 & \text{else} \end{cases}, \quad (4)$$

where T_k is the k -th task and P_j is the j -th platform.

Therefore, the allocation relationship R_{P-T} between a platform and a task can be expressed in terms of a $M \times N$ matrix.

$$R_{P-T} = [R_{P-T}(j, k)] = \begin{bmatrix} R_{P-T}(1,1) & R_{P-T}(1,2) & R_{P-T}(1,3) & \cdots & R_{P-T}(1,L) \\ R_{P-T}(2,1) & R_{P-T}(2,2) & R_{P-T}(2,3) & \cdots & R_{P-T}(2,L) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{P-T}(N,1) & R_{P-T}(N,2) & R_{P-T}(N,3) & \cdots & R_{P-T}(N,L) \end{bmatrix}_{N \times L} \quad (5)$$

For example, if there are 4 platforms and 3 tasks in the C2-organization. $N=4$ and $L=3$. T_1 requires P_1 and P_4 . T_2 needs P_1 and T_2 . T_3 requests P_2, P_3 , and P_4 .

Then R_{P-T} matrix is $R_{P-T} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}_{4 \times 3}$.

The command relationships between the decision-makers can be treated as a hierarchical and directional tree. It is composed of the decision-makers nodes (except root node) and the other decision-makers that have only one father decision-maker node. The expression for command relationships is given as:

$$R_{DM-DM}(m_1, m_2) = \begin{cases} 1 & \text{if the link between } DM_{m_1} \text{ and } DM_{m_2} \\ 0 & \text{else} \end{cases}, \quad (6)$$

where m_1 and m_2 are two decision-makers. The command relationship matrix is a square matrix that can be expressed as:

$$R_{DM-DM} = [R_{DM-DM}(m_1, m_2)] = \begin{bmatrix} R_{DM-DM}(1,1) & R_{DM-DM}(1,2) & \cdots & R_{DM-DM}(1,M) \\ R_{DM-DM}(2,1) & R_{DM-DM}(2,2) & \cdots & R_{DM-DM}(2,M) \\ \vdots & \vdots & \vdots & \vdots \\ R_{DM-DM}(M,1) & R_{DM-DM}(M,2) & \cdots & R_{DM-DM}(M,M) \end{bmatrix}_{M \times M} \quad (7)$$

When $m_1=m_2$, the self-relation always exists, but such relation is ignored in the paper. Therefore, the diagonal elements are always 0. As a hierarchical tree, there must be $(M-1)$ and only $(M-1)$ edges. For example, if there are 5 decision-makers within the system. $M=3$. DM_1 links to DM_2, DM_3 . The command relationships matrix can be expressed as $R_{DM-DM} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{3 \times 3}$.

Such directional relationship can be illustrated as single and double arrow lines (Fig. 1)

The execution relationship R_{DM-T}^E between decision-maker and task can be given as:

$$R_{DM-T}^E(m, i) = \begin{cases} 1 & \exists P_j, R_{DM-P}(m, j) = 1, R_{P-T}(j, i) = 1 \\ 0 & \text{else} \end{cases} \quad (8)$$

It means that one of the tasks of DM_m is T_i , if there is a P_j allocated to DM_m and assigned to T_i .

$$R_{DM-T}^E(i, k) = \begin{cases} 1 & \exists P_j, R_{DM-P}(i, j) = 1, R_{P-T}(j, k) = 1 \\ 0 & \text{else} \end{cases} \quad (9)$$

$$\begin{aligned}
 R_{DM-T}^E &= R_{DM-P} \times R_{P-T} \\
 &= [R_{DM-T}^E(i, k)] \\
 &= \begin{bmatrix} R_{DM-T}^E(1,1) & R_{DM-T}^E(1,2) & \cdots & R_{DM-T}^E(1,L) \\ R_{DM-T}^E(2,1) & R_{DM-T}^E(2,2) & \cdots & R_{DM-T}^E(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ R_{DM-T}^E(M,1) & R_{DM-T}^E(M,2) & \cdots & R_{DM-T}^E(M,L) \end{bmatrix}_{M \times L} \quad (10)
 \end{aligned}$$

For example, in this scenario, the R_{DM-T}^E is as follow under R_{DM-P} and R_{P-T}

$$R_{DM-T}^E = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

The command relationship R_{DM-T}^C among tasks is given as

$$R_{DM-T}^C(m, i) = \begin{cases} 1 \exists m \in DM, n \in DM^{T_i}, \\ \text{satisfy } \prod_{m \in DM^{T_i}} P(m, n) = 1, \\ \text{and } \forall m' \in \left\{ m' \mid \prod_{n \in DM^{T_i}} P(m', n) = 1 \cap m' \neq m \right\} \\ \text{satisfy } P(m, m') = 0 \\ 0, \quad \text{else} \end{cases} \quad (11)$$

$DM^{T_i} (i = 1, 2, 3, \dots, L)$ is the subset of decision-makers who involve in implementing tasks T_i . $DM_m^{T_i}$ is the element of subset DM^{T_i} . $DM_n^{T_i}$ is another element of subset DM^{T_i} . $P(m, n)$ is the directional connection that determines the matrix of decision-makers. It can be expressed as

$$P(m, n) = \begin{cases} 1 & \text{if existing a direct link between } DM_m^{T_i} \text{ and } DM_n^{T_i} \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned}
 R_{DM-T}^C &= P(R_{DM-DM}) \times R_{DM-T}^E \\
 &= [R_{DM-T}^C(i, k)] \\
 &= \begin{bmatrix} R_{DM-T}^E(1,1) & R_{DM-T}^E(1,2) & \cdots & R_{DM-T}^E(1,L) \\ R_{DM-T}^E(2,1) & R_{DM-T}^E(2,2) & \cdots & R_{DM-T}^E(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ R_{DM-T}^E(M,1) & R_{DM-T}^E(M,2) & \cdots & R_{DM-T}^E(M,L) \end{bmatrix}_{M \times L} \quad (12)
 \end{aligned}$$

where $P(R_{DM-DM})$ is the reachability matrix of R_{DM-DM} .

For example, in this scenario, the reachability matrix of R_{DM-DM} is itself. Hence, the R_{DM-T}^C is as follow under R_{DM-T}^E and R_{DM-DM}

$$R_{DM-T}^C = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

In a C2-organization system design, the system performance is highly related to its elements (platforms, decision-makers, and tasks) relations. How to adjust its elements to maximize the overall organizational effectiveness is one of the key issues. One of difficulties

is the dynamic uncertainty of mission environment. Man C2-organizations always cannot be well operated throughout the whole mission operation. A state-of-the-art organizational structure needs to be continuously updated and extended, which prompts our DODDS algorithm development. The following definitions are necessary.

Definition 1: The tasks implementation load of decision-makers DM_m in the task T_i can be expressed as:

$$DM_{mpm}^{T_i} = w_{mp} \cdot DT_i \cdot DI_i \cdot R_{DM-Tm}^E \quad (13)$$

where w_{mp} is the coefficient that accounts for the effect of tasks implementation load, DT_i is a durance or time required to accomplish task T_i (in terms of days or hours). DI_i is the strength of implementation load of task T_i . It is used to estimate how difficult the task T_i is. As essential and nature attributes of a task, both DT_i and DI_i are subjective. In the current study, we take the values from reference [23, 24].

Definition 2: The task collaboration load $DW_{colm}^{T_i}$ in task T_i . It can be expressed as:

$$DW_{colm}^{T_i} = w_{col} \cdot DT_i \cdot DI_i \cdot \sum_{m \neq m}^{N_{DM}} R_{DM-Tm}^E \cdot R_{DM-Tm}^E \quad (14)$$

where w_{col} is the coefficient that accounts for the task collaboration load.

Definition 3: The task command load $DW_{cmdm}^{T_i}$ in task T_i . It can be expressed as:

$$DW_{cmdm}^{T_i} = w_{cmd} \cdot DT_i \cdot DI_i \cdot R_{DM-Tm}^C \cdot \sum_{m \neq m}^{N_{DM}} R_{DM-Tm}^E \quad (15)$$

where w_{cmd} is the coefficient that accounts for the task command load.

Definition 4: The total decision work load $DW_m^{T_i}$ in task T_i accumulates all the above loads together. It can be expressed as:

$$DW_m^{T_i} = DW_{mpm}^{T_i} + DW_{colm}^{T_i} + DW_{cmdm}^{T_i} \quad (16)$$

Definition 5: The total task decision load DW^{T_i} in task T_i is the sum of all the decision working loads in the task T_i . It is given as:

$$DW^{T_i} = \sum_{DM_m^{T_i} \in DM^{T_i}} DW_m^{T_i} \quad (17)$$

where DM^{T_i} is the assembly of decision-makers to complete task T_i .

Definition 6: The decision contribution degree of decision-makers $DM_m^{T_i}$ in T_i is the ratio of total decision work load, $DW_m^{T_i}$ to the task decision load DW^{T_i} in task T_i . The ratio represents the importance of $DM_m^{T_i}$ in the task T_i 's decision work. It is denoted as $DG_m^{T_i}$ and expressed as:

$$DG_m^{T_i} = DW_m^{T_i} / DW^{T_i}. \tag{18}$$

Definition 7: The decision-makers DM_m 's decision work load $DW_m^{T_i}(\Delta T)$ in time-scale ΔT is the sum of decision work load of $DM_m^{T_i}$, where hereby $DM_m^{T_i}$ is the decision-maker that completes the task within ΔT . $DW_m^{T_i}(\Delta T)$ can be expressed as:

$$DW_m^{T_i}(\Delta T) = \sum_{i \in T} DW_m^{T_i} \cdot \delta_{\Delta T}^{T_i}, \tag{19}$$

where the $\Delta T = [t_{LB}, t_{UB}]$ is a time interval. t_{LB} and t_{UB} stand for the starting time and ending time. $\delta_{\Delta T}^{T_i}$ is a sign function judging whether task T_i is executed within ΔT or not. The sign function can be expressed as:

$$\delta_{\Delta T}^{T_i} = \begin{cases} 1 & T_i \text{ executed within time-scale } \Delta T \\ 0 & \text{else} \end{cases}. \tag{20}$$

Definition 8: The quality of decision-making $q_m(\Delta T)$ can be calculated based on the ratio of quality of decision-makers $DM_m^{T_i}$ within ΔT , λ . The ratio is defined to measure the effectiveness required to complete the decision-making by $DM_m^{T_i}$ within ΔT . It depends on the $DM_m^{T_i}$'s decision work load $DW_m(\Delta T)$ and the upper limit of the working load within ΔT , DWB_m . It is expressed as:

$$\lambda = (DW_m(\Delta T) - DWB_m) / DWB_m. \tag{21}$$

It can be called loading capacity ratio of decision-makers. Its value is between 0 and 1, the value of $DW_m(\Delta T)$ must be less than DWB_m limit. As $DW_m(\Delta T)$ exceeds that limit, it cannot provide adequate decision-making capability. Usually $DM_m^{T_i}$ have a certain robustness to withstand a certain degree of over-load.

If the ratio λ is small, the gradient of decline in the quality of the work decision-making is relatively slow, and the value is close to unity. If λ takes a medium, in rang form 0.3 to 0.7, the decision-making quality of $DM_m^{T_i}$ declines rapidly. If λ becomes larger, the decision work load of $DM_m^{T_i}$ is far beyond its up-load.

The quality of the work of decision-making is very low and even reaches to zero.

Therefore, we define the quality of decision-making $q_m(\Delta T)$ as:

$$q_m(\Delta T) = \begin{cases} 1 & \lambda < 0 \\ \frac{\exp^{10(1-2\lambda)}}{1 + \exp^{10(1-2\lambda)}} & \lambda \geq 0 \end{cases} \tag{22}$$

Definition 9: The task's quality of decision-making $DQ^{T_i}(\Delta T)$ is defined as the summation of all the quality of decision-making participated in the task pool. It can be calculated as:

$$DQ^{T_i}(\Delta T) = \sum_{DM_m^{T_i} \in DM^{T_i}} DG_m^{T_i} \cdot q_m(\Delta T) \cdot DA_m(\Delta T) \cdot \delta_{\Delta T}^{T_i}. \tag{23}$$

This value represents the total quality of accomplished decision-making work.

Definition 10: The temporal mean value of all quality of decision-making tasks, $DC(\Delta T)$, represents the organizations' effective measure of decision-makings. It estimates the whole capability of C2 organizational decision making system.

Given the ability of C2 organizational decision making within ΔT , $DC(\Delta T)$ can be given as:

$$DC(\Delta T) = \sum_{T_i \in T(\Delta T)} \frac{\text{Min}(t_{UB}, t_{f_i}) - \text{Max}(t_{LB}, t_{s_i})}{|\Delta T|} \cdot \frac{DQ^{T_i}(\Delta T)}{|T(\Delta T)|}, \tag{24}$$

where t_{s_i} and t_{f_i} stand for the starting time and ending time, respectively. $|\Delta T|$ is a period time required for a task executing within ΔT . There is some idle period of time in the time domain ΔT due to no task executed during the period. It can be ignored since it is not related to the organization decision-making capacity. $T(\Delta T)$ is the tasks assembly within ΔT . The numerator $\text{Min}(t_{UB}, t_{f_i}) - \text{Max}(t_{LB}, t_{s_i})$ refers the time length of executing task T_i within ΔT . The ratio of $[\text{Min}(t_{UB}, t_{f_i}) - \text{Max}(t_{LB}, t_{s_i})] / |\Delta T|$ occupies significant part of computation for $DC(H)$.

When $t_{UB} \rightarrow t_{LB}$, $|\Delta T| = 0$, ΔT represents the transient time at t_{LB} . Therefore, one has:

$$\lim_{t_{UB} \rightarrow t_{LB}} DC(\Delta T) = \sum_{T_i \in T(t_{LB})} \frac{DQ^{T_i}(t_{LB})}{|T(t_{LB})|}. \tag{25}$$

At this point $DC(\Delta T)$ can be signed as $DC(t_{LB})$ denoted the organization decision-making capacity at the transient time at t_{LB} . The equation above can be understood that there is a major task during that period in the context of a period of the time domain. The tasks become important at that moment. If there is main task, all tasks could reach to the equal level of importance.

3 Dynamic optimization of decision-maker and decision-layer structure (DODDS) model

The proposed DODDS model considers both the adjustable costs of the structure and dynamic decision-making capacity. The details of model algorithm are following.

Definition 11: The adjustable costs of the structure (AC) is defined as the restructure costs due to the changes of command-relationships between decision-makers and of the control-relationships between decision-makers and platforms.

Suppose that the decision-layer-structure is G_{DLS} and optimized value is G'_{DLS} . The adjustable costs $AC(G_{DLS}, G'_{DLS})$ can be expressed as follows:

$$AC(G_{DLS}, G'_{DLS}) = \left(\frac{W_D}{N_{DM}}\right) \cdot \sum_{m_2=1}^{N_{DM}} \sum_{m_1=1}^{N_{DM}} |R'_{DM-DM}(m_1, m_2) - R_{DM-DM}(m_1, m_2)| + \frac{W_p}{2N_p} \cdot \sum_{j=1}^{N_p} \sum_{m=1}^{N_{DM}} |R'_{DM-p}(m, j) - R_{DM-p}(m, j)| \quad (26)$$

where the first term accounts for the amount of the changed command links between decision-makers, and the second term represents the amount of the changed control links between decision-makers and platforms. W_D and W_p represent the costs per change in the command and control links respectively. Since there are some direct links, the command-to-change cannot be taken into consideration, repeatedly. N_{DM} and N_p are the total numbers of decision-makers and platforms, respectively. Obviously ACs depends on the organization size. For example, if W_D or W_p remain constant, the larger organization is, the smaller the cost. Therefore, the proportions of the changes become important in evaluating the adjustable costs.

Definition 12: Decision-making capacity (DMC) is defined as the difference of between the organizational-decision-making-capacity and optimal one. According to Definition 10, it can be expressed as:

$$DMC(G_{DLS}, G'_{DLS}, \Delta T) = DC(G'_{DLS}, \Delta T) - DC(G_{DLS}, \Delta T) \quad (27)$$

where $DC(G_{DLS}, \Delta T)$ and $DC(G'_{DLS}, \Delta T)$ stand for the organizational-decision-making-capacities of the decision-layer-structures G_{DLS} and optimal value G'_{DLS} respectively.

Definition 13: The dynamic optimization of decision-layer-structure (DLC) gives the profit for an organization. It can be expressed as:

$$DLC(G_{DLS}, G'_{DLS}, \Delta T) = W^P \cdot DMC(G_{DLS}, G'_{DLS}, \Delta T) - W^R \cdot AC(G_{DLS}, G'_{DLS}, \Delta T) \quad (28)$$

where W^P and W^R are the weights of organizational-decision-making-capacity and the adjustable costs respectively. The smaller the W^R , the more flexible the organization is and the less dynamic adjustable costs. The case $DLC(G_{DLS}, G'_{DLS}, \Delta T) > 0$ implies two facts. One is that the organizational-decision-making-capacity is greater than the adjustable costs. The other is that the organizational decision-making can be improved by using utilizing properly adaptive optimizations.

There exit some constrains [18] that should be satisfied during the dynamic optimization. These constrain conditions (restriction) are as follows.

The restriction on the relationship between decision-makers and platforms:

$$\sum_{m=1}^{N_{DM}} R'_{DM-p}(m, j) = 1, j = 1, 2, \dots, N_p \quad (29)$$

The restriction on the upper limit of a decision maker controlled platforms:

$$\sum_{m=1}^{N_p} R'_{DM-p}(m, j) \leq \overline{CN}, m = 1, 2, \dots, N_{DM} \quad (30)$$

The restriction on the decision tree that has an unique root must satisfy the following one:

$$\sum_{m \neq m^*}^{N_{DM}} R'_{DM-DM}(m, m^*) = 0 \quad (31)$$

Any nodes except the root has a single parent node:

$$\forall m' \neq m^*, \sum_{m \neq m'}^{N_{DM}} R'_{DM-DM}(m, m') = 1 \quad (32)$$

From presentation and definitions above, the final DODDS optimization model is summarized as:

$$\begin{aligned} &\max DLC(G_{DLS}, G'_{DLS}, \Delta T) \\ &\left\{ \begin{aligned} &G'_{DLS} = (R'_{DM-p}, R'_{DM-DM}) \\ &\sum_{m=1}^{N_{DM}} R'_{DM-p}(m, j) = 1, j = 1, 2, \dots, N_p \\ &\sum_{j=1}^{N_p} R'_{DM-p}(m, j) \leq \overline{CN}, m = 1, 2, \dots, N_{DM} \\ &\sum_{m \neq m^*}^{N_{DM}} R'_{DM-DM}(m, m^*) = 0 \\ &\sum_{m \neq m'}^{N_{DM}} R'_{DM-DM}(m, m') = 1 \\ &R'_{DM-p}(m, j) \in \{0, 1\}, m = 1, 2, \dots, N_{DM}, j = 1, 2, \dots, N_p \\ &R'_{DM-DM}(m_1, m_2) \in \{0, 1\}, m_1, m_2 = 1, 2, \dots, N_{DM} \end{aligned} \right. \quad (33) \end{aligned}$$

4 Conclusions and future works

This paper presents a model of Decision-makers and Dynamic Optimization of Decision-layer Structure (DODDS) for a C2-structured organization in operational SoS. Through intensive and extensive literature studies and the empirical experiments, we believe that the performance of C2- structured organization is always much lower than expected value because of environmental uncertainty and system changes. To address the issue, we extended the exiting model of C2




decision-layer-structure by introducing the time-domain, by employing the NISA and horizon (time-based) discretization. The model concepts are addressed, terminology are defined, and the optimization objective function of the model with constrains are derived. The model will be implemented in several scenario tests to be presented in the coming paper to demonstrate the model feasibility and applicability.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant number 91024006, 61273322, 71001105 and 70771109.

References

- [1] Thomas A, Turner T, Soderlund S 2009 Net-Centric Adapter for Legacy Systems *IEEE Systems Journal* **3**(3) 336-342
- [2] Butterfield M L, Pearlman J S 2008 A System-of-Systems Engineering GEOSS: Architectural Approach *IEEE Systems Journal* **2**(3) 321-332
- [3] Liu Z, Yang D S, Wen D, Zhang W M, Mao W J 2011 Cyber-Physical-Social Systems for Command and Control *IEEE Intelligent Systems* **26**(4) 92–96
- [4] Qian M, Liu Z, Wang J, Yao L, Zhang W M 2010 Coordination-theoretic Approach to Modelling Grid Service Composition Process *Journal of Systems Engineering and Electronics* **21**(4) 713-720
- [5] Carley K M 1995 Computational and Mathematical Organization Theory: Perspective and Directions *Computational and Mathematical Organization Theory* **1**(1) 39-56
- [6] Rajan R G, Wulf J 2006 The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies *Review of Economics and Statistics* **88**(4) 759-773
- [7] Chuma H 2006 Increasing Complexity and Limits of Organization in The Microlithography Industry: Implications for Japanese Science-based Industries *Res Policy* **35** 393-411
- [8] Barr J, Nobuyuki H 2008 Organizations Undertaking Complex Projects in Uncertain Environments *Journal of Economic Interaction and Coordination* **3**(2) 119-135
- [9] Lin Z 2006 Environmental Determination or Organizational Design: An Exploration of Organizational Decision Making Under Environmental Uncertainty Simulation *Modelling Practice and Theory* **14**(4) 438-453
- [10] Carley K M, Prietula M J 1994 *Computational Organization Theory* Lawrence Erlbaum Associates
- [11] Carley K M, Lin Z 1995 Organizational Designs Suited to High Performance under Stress *IEEE Transactions on Systems Man and Cybernetics* **25**(1) 221-230
- [12] Carley K M, Svoboda D M 1996 Modelling Organizational Adaptation as a Simulated Annealing Process *Sociological Methods and Research* **25**(1) 138-168
- [13] Krackhardt D, Carley K M 1998 A PCANS Model of Structure in Organization *Proceedings of the 1998 International Symposium on Command and Control Research and Technology* 113-119
- [14] Levchuk G M, Levchuk Y N, Luo J, Pattipati K R, Kleinman D L 2002 Normative Design of Organizations I: Mission Planning *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **32**(3) 346-359
- [15] Levchuk G M, Levchuk Y N, Luo J, Pattipati K R, Kleinman D L 2002 Normative Design of Organizations II: Organizational Structure *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, **32**(3):360-375
- [16] Levchuk G M, Levchuk Y N, Luo J, Pattipati K R and Kleinman D L 2004 Normative Design of Project-based Organizations-Part III: Modelling Congruent, Robust, and Adaptive Organizations *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **34**(3) 337-350
- [17] Meirina C, Levchuk Y N, Levchuk G M, Pattipati K R 2008 A Markov Decision Problem Approach to Goal Attainment *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **38**(1) 116-132

Authors	
	<p>Yanghe Feng</p> <p>University studies: Master degree in engineering from National University of Defense Technology, China, in 2009. He is currently pursuing the Ph.D. degree with Key Laboratory for Information System Engineering, National University of Defense Technology.</p> <p>Scientific interest: data mining, social computing and complex networks.</p>
	<p>Dr. Baoxin Xiu, born in 1977</p> <p>Current position, grades: Ph.D. and an associate professor in Information Systems Engineering Laboratory at NUDT.</p> <p>Scientific interest: granular computing, computational and mathematical organization theory and social computing.</p>
	<p>Dr. Zhong Liu, born in 1968</p> <p>Current position, grades: Professor at NUDT and the deputy director of the Science and Technology in Information Systems Engineering Laboratory at NUDT.</p> <p>University studies: Ph.D. degree in management science and engineering from the National University of Defense Technology (NUDT) in 2000.</p> <p>Scientific interest: command and control organization, planning systems, and intelligent systems.</p>

Research on the method of determining the service scope of civil airport – the case of Chengdu Shuangliu airport

Zhao Suxia^{1, 2*}, Yang Xuemei², Sun Peihong³

¹ *School of Transportation and Logistics, Southwest Jiaotong University, No.111 of Northbound Section 1, Second Ring Rd., Chengdu, 610031, Sichuan, China*

² *Sichuan Vocational and Technical College of Communications, Chengdu, 611130, Sichuan, China*

³ *Twenty First Century Aerospace Technology Co., Ltd. Beijing, 100096, China*

Received 1 March 2014, www.tsi.lv

Abstract

Airport hinterland is the most important of influencing factors for civil aviation airport layout. By using methods of breaking point, expert consultation and entropy weights, the formula of computing the acting radius for airport hinterland was set up according to breaking point theory, and the idea and method were proposed to determine the service scope of an airport in an economic region. Then demonstrate the rationality of airport layout in the Chengdu-Chongqing economic zone by calculating the actual case of airport in the Chengdu-Chongqing economic zone. And get a conclusion that Leshan airport is inappropriate expansion in the near future.

Keywords: Civil airport; Breaking point theory; Entropy weights; Service scope

1 Introduction

Because of several constraints such as environment, resources and land use, major city groups in the world have a common problem of high density airports, the use of core city airport resource has been saturated, flight delay are serious and airport service quality has declined as well, but the airport facilities of surrounding cities are idle and lots of airport resources have not been fully used.

Civil airport lives with city. It is a crucial place for exchanging information between cities and cities and people communication, is also an important basic platform for regional economic social development. With the development trend of civil airport towards large-scale and hub type, including from regional city group to small towns around airport. The function coverage of some civil airports not only serves its home city also gradually extend to serve the entire economic region. Therefore, to accurately define the area of civil airport ground service, carry out overall planning in the region, and reasonably distribute aviation resources, avoid duplicated construction are important for the sustainable development of transportation system in the regional economy.

Airport service area or airport radiation circle means the area that airport ground radiation capability can cover. It includes the roads, railways and other ground transportation connected to the airport, which may generate passengers and goods supply for airport, and make the passengers and export goods consumption through airport.

According to the features of service needs, there are two levels of meaning should be taken into consideration. The first level is airport space radiation, i.e. radius from the airport; the second level is airport economic radiation, including core market and radiation market. The civil airport service objects studied in this paper are the passengers who need land to air transfer inside airport, airport ground radiation area is the airport ground attractive area under the background of integrated transport system and regional economic development.

In recent years, some experts and scholars using quantitative methods to study the reasonableness evaluation for the choice of airport site, mostly considering the factors of owned by city with technical point of view to study without the radiation factors around the airport; Zhang Xiaobin, Wang Wuke [1] have deeply analysed other factors affecting the construction of airport which can be divided into two categories; the background reflecting the development potential and service -based radius reflecting the potential pressure. For those have good economic foundation, but outside the airport layout that cannot meet the needs of small and medium cities, the recent urban planning should consider reserving the area for airport. Yang Xiuyun, Yao Shujie [2] thought that the structure, ground transportation and economic openness region are also major factors affecting the development of the airport and should be included in the analysis model. When other conditions remain unchanged, the impact of regional location on the airport development is significant. Wang Jiao, Jin Fengjun [3] using two different methods of the shortest

* Corresponding author - Tel: +86-13982284708; fax: +86-028-82680899; E-mail: 10103288@qq.com

distance and reasonable service radius to divide the space radiation range of each airport. With the rapid development of regional economy, integrated transport, it seems that both methods have some limitations today.

Based on the absorption of results from previous studies, through expert consultation analysis, the airport ground radiation impact factor is selected and uses Entropy Method and combination weight to obtain the relative attractiveness M of the airport. Then the airport ground services area is qualitatively and quantitatively determined by using breaking point theory to identify the airport ground radiation range which could guide the urban airport planning and construction within the economic area.

2 The definition of the civil airport attraction

2.1 THE ESTABLISHMENT OF EVALUATION INDEX SYSTEM OF CIVIL AIRPORT ATTRACTION

The factors which affect ground radiation scope of civil airport include the following items [4].

The first point is economic attractiveness of the hinterland city. Economic attractiveness is the major indicator which reflects degree of the city's production factors exchange with outsides, including resource distribution, industrial scale, the number of the employed population, economic structure, the openness of city, etc. These factors will affect the generation and attracting passenger flow, and determine the demand characteristics of civil airports.

The second point is accessibility and smoothing of transportation network around airport. The development of civil airport depends on the support of transportation network. In general, the wider of airport services cover, the farther the average distance to the airport, the weaker accessibility is, and vice versa; the greater population of the airport serves, indicating that more people use is, and the weaker smoothing is.

The third is overall strength of the airport. The overall strength of the airport is core competitiveness to attract the passenger and cargo traffic, and it is the most important factor in competition with the surrounding airports. In theory, the airport's overall strength includes airports modern equipment, management techniques, sufficient routes resources, unified and efficient operational means organizational skills, which are embodies the strength of the airport itself, and affects the degree of aggregation passenger.

Based on above factors and widely soliciting opinions from industry experts, in the principle of indicators selected scientific, operability and the combination of qualitative and quantitative indicators, the civil airport attractiveness indicator system as table1 is built.

TABLE 1 Civil airport attractiveness indicator system

Sub-function	Criterion	Indicators	Notes
Hinterland city attractiveness	Socio-economic factor	Openness of city	quantitative
		Per capita disposable income	quantitative
		GDP	quantitative
Ground transportation network	The ground transportation system	Proportion of the tertiary industry	quantitative
		Ground transportation accessibility	qualitative
Airport overall strength	Air route resource	Ground transportation smoothing	qualitative
		Air routes number	quantitative
	Transport organization ability	Navigation mileage	quantitative
		Aircraft movement	quantitative
		Airport cargo and passenger throughput	quantitative
		Transfer efficiency	qualitative

The access methods to specific qualitative or quantitative indicators are explained below.

(1) Openness of the city

Openness of the city mainly refers to the international exchange and cooperation and various factors of production flow inside and outside the city, which covers the areas of trade, investment, services, human resources, etc., it is an important indicator of the attractiveness of the city, and detailed evaluation methods can be seen in the literature [5].

(2) Per capita disposable income and GDP

These two factors reflect overall economic strength of the city and its local residents. They are important indicators for one city attractiveness.

(3) The proportion of the tertiary industry

The proportion of the tertiary industry reflects the degree of economic development of countries or regions, and there is a positive correlation relationship between it and air transport demand.

(4) Ground transportation network

Well-developed transportation system is an important manifestation of the airport's overall strength, and it is also an important reference index for passengers to select airport. Smoothing and accessibility are two core indicators of the airport transportation system, and smoothing affects the attractiveness for people around the airport and reflects the airport efficiency on integrated transport services; while accessibility embodies airport's universal service function and reflects the possibility transport services for airport passengers. Expert consultation is applied for these two evaluation methods.

(5) Route number and navigable mileage

Routes refer to the navigation lines for aircraft from one place (the starting point) fly to another place (end), and navigable mileage is the total length of routes. They are important statistical indicators to reflect the extent of route network expansion for an air transport enterprise production and business activities. Its statistical method can be seen in literature [6].

(6) Aircraft movements & airport cargo and passenger throughput

Aircraft movement refers to the maximum allowable vehicles on the airport runway, during a certain period,

and the general statistical indicators include annual average, daily average and daily peak hour aircraft movements. The number of vehicles in annual average is used here.

Since the airport's attractiveness is mostly reflected on passenger demand, and airport passenger throughput is selected as indicator here. Airport Passenger Throughput refers to the number of airport passengers arrive and depart during the reporting period which is computed in terms of Person. To some extent it reflects the size of the airport's handling capacity.

(7) Transfer efficiency

It refers to efficiency for the passengers to leave the airport or at the airport transit. It reflects the airport's internal management and organizational capability. This indicator is evaluated with the Delphi method.

2.2 NON-DIMENSION OF INDICATORS

As different indicators have different units resulting into lack of incommensurability, normalizing process for indicator evaluating data including positive, moderate, reverse and interval indicator is necessary. The characteristics of the data will be reflected by this process and the process method can be referred to literature [7].

2.3 DETERMINING THE AIRPORT ATTRACTION BY ENTROPY METHOD AND COMBINATION WEIGHTING METHOD

2.3.1 Calculation steps of entropy method

Entropy was once a concept of thermodynamics. At first it was introduced to information theory by c. e. Shannon and was called information entropy. The information entropy is an independent of the thermodynamic entropy concept, but has the basic properties of thermodynamic entropy, and has more extensive and universal significance, known as generalized entropy. The entropy weight method is a kind of objective empowerment. In the process of concrete use, based on the variation degree of each index, the entropy weight method will make use of information entropy to calculate the entropy weight of each index by entropy, then modify the weight of each index by entropy, by which more objective index weight will be obtained.

The specific calculation steps are as follows:

Step1. Through the field survey data and expert consultation on N evaluation objects and M index evaluation, the object and the index of evaluation matrix $R = (r_{ij})_{m \times n}$ will be formed.

Step2. According to the definition of entropy, in M index evaluation of N evaluated objects, the entropy of the i^{th} indicators will be defined as:

$$H_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij}, \text{ where } f_{ij} = \frac{r_{ij}}{\sum_{j=1}^n r_{ij}},$$

$$k = 1/\ln n, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_0.$$

Step3. To calculate weight of the indicators, the entropy weight of the i^{th} indicators will be defined as: $w_i = (1 - H_i) / (m - \sum_{i=1}^m H_i)$.

Step4. To analyse entropy and entropy weight, when all evaluation objects is the same in a certain index, if the entropy value is 1 and the entropy is 0, it indicates that the index fails to provide useful information and may be cancelled.

2.3.2 Calculating the steps by combination weighting method as follows

Combination weighting method combines expert consultation method, analytic hierarchy process (AHP) and entropy weight method. The method not only considers the actual situation of the problem, but also combines with implicit knowledge of the experts in this field. It will avoid subjective method of assumptions and depend on the original data of mechanical properties too much by which focused analyses makes it explicit. It is more suitable for the related research work in complex system in social-economic field. The specific calculation steps are as follows.

Step1. The resulting weight by expert consultation method and analytic hierarchy process is: $W_1 = (w_1^1, w_1^2, \dots, w_1^k)$.

Step2. The resulting weight by Entropy weight method is $W_2 = (w_2^1, w_2^2, \dots, w_2^k)$.

Step3. To make $w_0^i = w_1^i \times w_2^i$ a group of weight vector $W_0 = (w_0^1, w_0^2, \dots, w_0^k)$ will be obtained.

Step4. In order to satisfy the normalization of weight, weight vector will be standardized $w^i = w_0^i / \sum_{i=1}^k w_0^i$, so combination weight vector is $W = (w^1, w^2, \dots, w^k)$.

3 To confirm service scope of civil airport

The breaking point theory is about urban and regional interaction. The theory argues that a city's attractiveness to the surrounding area is proportional to the distance, and it is inversely proportional to the square of the distance. So boundary points (i.e. breaking point) of the affect areas by two cities are shown by formula such as 3-1 show.

The airport can effect deeply on the development of adjacent area. This effect also varies due to the scale of the airport and may weaken gradually with the increase of distance. That is to say, there is also a distance attenuation rule between airport and distant radiation. So radiation force of two adjacent airports will reach a balance and form a balancing point. Therefore, the calculation model of breaking point for airport will be obtained.

There must be several airports in an economic area and they will compete. Passengers choose one of them

because of its appeal. So in this article we want to build the improved model of weighted breaking point to evaluate comprehensively charm index for appealing passengers. The attracting index will be obtained by calculating synthetically index system of gravity at the airport in formula (1) by using the entropy weight method and combination weighting method. The resulting fitting curve of the breaking point will act as the ground space radiation of the research object.

$$D_a = \frac{D_{ab}}{1 + \sqrt{M_b / M_a}}, \tag{1}$$

where D_a is the distance between the breaking point and civil airport A. D_{ab} is the distance between two competitive airport A and B. $M_b = \sum_{i=1}^n \omega_i Y_{bi}$, $M_a = \sum_{i=1}^n \omega_i Y_{ai}$, where Y_{ai} and Y_{bi} are influenced factors respectively. I will normalize the effect to airport A and B. N is the number of factors.

We will obtain the breaking point D_{ab} by substituting gravity results of the civil airport calculated by integrated weight method and quantitative measured data into formula (1). Therefore, the obtained breaking point will be important reference point for determining service scope, not those simple connecting breaking points. So the location of the breaking pointing between target airport and other airports in this area will be calculated by this model. If all the breaking points are obtained, service scope of the target airport will be determined by closed curve fitting formulated by MATLAB software.

4 Application case

For example, in Chengdu-Chongqing economic zone, according to airport business ranking in 2011, the top four airports are selected. Chengdu Shuangliu Airport is marked as A, Chongqing Jiangbei Airport is marked as B, Mianyang Nanjiao Airport is marked as C and Yibin Wuliangye Airport is marked as D. Passenger throughput of these four airports is (2.9, 1.9, 0.06, 0.03)×107 respectively in 2011. In view of the coordination and cohesion between airport and road network, their distance will be marked with urban road minimum distance. It is shown in figure 1. (Unit: Km)

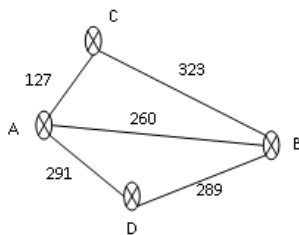


FIGURE 1 The top four airports network topology in Chengdu-Chongqing economic zone

4.1 DETERMINE THE BREAKING POINT POSITION OF AIRPORT A AND COMPETITOR

Determine the breaking point position of Airport A and competitor in Figure1 by calculation tool ‘Yaahp’ (Yet another AHP), entropy weight method and portfolio weighting method according to breaking point formula. The detailed process is followed.

Step1. According to questionnaire on experts, through pairwise relationship on every factor, we can obtain effect weight of airport attraction caused by every index of table 2 by ‘Yaahp’, as shown in table 2.

TABLE 2 Index weight by expert consultation method and AHP

Sub-function	Criterion	Indicators
Hinterland city attractiveness 0.26	Socio-economic factor 0.26	1. Openness of city 0.09
		2. Per capita disposable income 0.08
Ground transportation network 0.25	The ground transportation system 0.25	3. GDP 0.06
		4. The proportion of the tertiary industry 0.03
		5. Ground transportation accessibility 0.13
Airport overall strength 0.49	Air route network 0.20	6. Ground transportation smoothing 0.12
		7. Air routes number 0.12
	Transport organization ability 0.29	8. Navigation mileage 0.08
		9. Aircraft movement 0.09
		10. Airport cargo and passenger throughput 0.12
		11. Transfer efficiency 0.08

Step2. Apply dimensionless method to the practical data which is collected from all airports and some empirical index from experts’ scoring and then obtain the evaluation matrix R.

$$R = \begin{bmatrix} 0.54 & 0.41 & 0.02 & 0.03 \\ 0.27 & 0.25 & 0.26 & 0.22 \\ 0.43 & 0.47 & 0.05 & 0.04 \\ 0.33 & 0.51 & 0.08 & 0.08 \\ 0.26 & 0.28 & 0.23 & 0.23 \\ 0.24 & 0.29 & 0.24 & 0.24 \\ 0.41 & 0.34 & 0.14 & 0.12 \\ 0.38 & 0.33 & 0.15 & 0.14 \\ 0.52 & 0.33 & 0.07 & 0.06 \\ 0.43 & 0.34 & 0.13 & 0.11 \\ 0.38 & 0.35 & 0.14 & 0.13 \end{bmatrix}$$

Step3. According to calculate steps of entropy method in section 2.3.1, calculate the attraction indexes weight at the airport shown in table 3.

Step4. The attraction index of airport calculated through combination weighting method in section 2.3.1 was shown in table 4.

TABLE 3 Airport attraction indexes weight calculated by entropy method

	Openness of city	Per capita disposable income	GDP	Proportion of the tertiary industry	Ground transportation accessibility	Ground transportation smoothing	Air routes number	Navigation mileage	Aircraft Movement	Airport cargo and passenger throughput	Transfer efficiency
Weight	0.267	0.00	0.196	0.157	0.000	0.008	0.069	0.046	0.184	0.007	0.059
	564	137	655	229	701	815	882	409	253	625	498

TABLE4 Airport comprehensive attraction index weight

	Openness of city	Per capita disposable income	GDP	Proportion of the tertiary industry	Ground transportation accessibility	Ground transportation smoothing	Air routes number	Navigation mileage	Aircraft Movement	Airport cargo and passenger throughput	Transfer efficiency
Weight	0.3293	0.0015	0.1613	0.0645	0.0012	0.0145	0.1147	0.0508	0.2268	0.0125	0.0651

Step5. The breaking point position of airport A and competitor

To combine and calculate quantitative matrix R and acquired index weight, we will see that attraction of airports A, B, C and D are respectively 3.278, 2.867, 0.568, 0.473. Then generate the obtained data into the formula (1), the breaking point position of Airport A will form. By matching we will see influencing range shown in figure2.

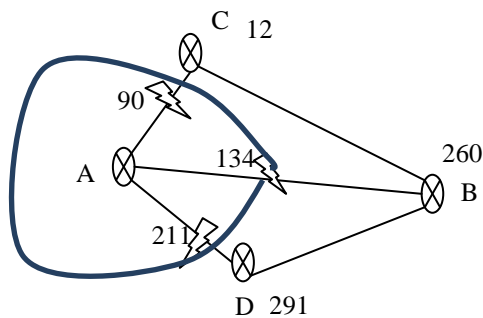


FIGURE 2 The breaking point position and influencing scope of airport A in this zone



FIGURE 3 Breaking point of Chengdu Shuangliu Airport

4.2 THINKING ABOUT PLANNING AND CONSTRUCTION OF THE AIRPORTS IN CHENGDU-CHONGQING ECONOMIC ZONE

According to breaking point and influencing range of Chengdu Shuangliu Airport acquired just now, shown in

figure 3, we can understand Leshan, about 169 km from Chengdu, is in this influencing range. Under development environment of the comprehensive transportation system, decision makers should apply policy guidance and management from the macroscopic angle. What is more, due to the development of highway and high-speed rail and the cause of airport resource around and self-position of airport attraction, Leshan Airport shouldn't be expanded aimlessly in the near future. The related institutions should invest auxiliary transportation infrastructure of airport, so as to realize optimization of system capacity and service in Chengdu-Chongqing economic zone airport.

Because service range of different airports crosses each other and competes drastically, in constructing and operating at airport, the relevant departments firstly consider regional coordinated development and apply flexible guidelines for the development strategy and development goals each airport. According to policy, actual condition at airport and development strategy of the airline, dynamically adjusting will be applied for development strategy of the airport. At the same time, by means of price demand control and so on, apply dynamic adjustment distribution of traffic volume in order to guarantee the maximum use of resources at the airport.

5 Conclusions

Combined with the factors that affect ground service range of the regional civil aviation airport, the formulas of calculating ground service scope of the civil aviation airport will be put forward. Among these factors, we consider not only airport itself and the surrounding geographical characteristics but also the layout of surrounding airport. Then calculating by model, we can solve ground service scope of the civil aviation airport, determine location selection and reasonable structure and avoid repetitive construction and malignant competition. And also, aiming at some blind spots on reasonable expanding in western zone especially those with plentiful travel resources, we will establish suitable layout method for China's national conditions with airport ground service theory.




Due to complicated factors involved, the theory may lack of Consideration of factors such as passenger preference. We will improve it in the future.

Acknowledgments

This research was supported in part by Grant No. 2012ZR0168 from Science & Technology Department of Sichuan Province.

References

- [1] Xiaobin Zhang, Wuke Wang 2010 The Research of Small and Medium-sized Airport Layout under the Perspective of Urban Planning *The planning and innovation: the annual conference on Chinese urban planning in 2010, Oct.15* 310-318 Chongqing, China: Urban Planning Society of China
- [2] Xiuyun Yang, Shujie Yao 2009 The determinants of airport industry development in China: empirical analysis based on augmented production function *Modern Economic Science* 31(1) 43-45 DOI: 1002-2848-2009(01)-0042-08
- [3] Jiao'e Wang, Fengjun Jin, Wei Sun, Teqi Dai, Chengjun Wang 2006 Research on spatial distribution and service level of Chinese airport system *Acta Geographica Sinica* 61(8) 829-838
- [4] Li Zhen Zhao 2006 The influencing factors and development conditions of integrated transport hub *Coordinated Transportation* 330(6) 9-11
- [5] Fa Ming Wang 2008 The framework of the evaluation index system of opening up: take hangzhou as example. *Reform* 175(9) 51-57
- [6] Zhipo Zhang 1994 An evaluation on statistics of air routes and mileage *CAET* 146(2) 31-32
- [7] Yaorong Cheng, Huayi San, Fenggen Liu 2008 Acting sphere determination of logistics park and allocation calculation of logistics volume *Journal of Traffic and Transportation Engineering* 8(6) 123 DOI: 167121637(2008)0620122205
- [8] Bonnefoy P A 2006 *Emergence of secondary airports and dynamics of regional airport systems in the United States USA*: Massachusetts Institute of Technology
- [9] Chongjun Xiong 2006 *Several coordinated development problems research on comprehensive transportation in China* Nanjing, China: Nanjing University of Aeronautics and Astronautics Doctoral Dissertation (1028709 06-0017)

Authors	
	<p>Suxia Zhao, born in October 23, 1981, China</p> <p>Current position, grades: Lecturer of transportation systems engineering in Sichuan Vocational and Technical College of Communications. University studies: Dalian jiaotong university, safety engineering (Bachelor); Southwest Jiaotong University, safety technology and engineering (Master); A PhD student major in transportation systems engineering in Southwest Jiaotong University now. Scientific interest: Airport planning and management, transportation systems engineering. Publications: 10 papers Experience: Since 2008 she has been working as a PhD student in Southwest Jiaotong University, during this period, she turned to the study of airport planning, and got support from Science & Technology Department of Sichuan Province, and published some papers on airport planning.</p>
	<p>Xuemei Yang, born in January 12, 1975, China</p> <p>Current position, grades: Lecture of Department of Economic Management in Sichuan Vocational and Technical college of Communications University studies: Southwest University of Finance and Economics (Master); Sichuan International Studies University (Bachelor) Scientific interest: Regional economic Publications: 2 papers Experience: After graduation from college, she has been working as a teacher in Sichuan Vocational and Technical college of Communications, and get love from her students.</p>
	<p>Peihong Sun, born in Mar 27, 1983, China</p> <p>Current position, grades: Project Manager, Twenty First Century Aerospace Technology Co., Ltd. University studies: Beijing University of Aeronautics & Astronautics, Aerospace engineering (Master); Beijing Normal University, Electronic Information Science and Technology (Bachelor) Scientific interest: GIS Publications: 3 papers and 2 patents Experience: After graduation from college, she has been working as an information technology worker in worker in Twenty First Century Aerospace Technology Co., Ltd. Two years ago, she became a project Manager, and obtain 2 patents on information communication</p>

Tourist flow-control study based on utility

Liu Zhusheng*

Business School, Sichuan University, Wangjiang Road 29, Chengdu 610064, China

Received 1 January 2014, www.tsi.lv

Abstract

The key to solving tourist congestion during peak seasons is to improve the level of management and control the tourist flow in a scientific manner. Scenic capacity is usually set to a fixed value. Consequently, many plans have been developed, but have often produced poor results in practice. The utility concept in economics was introduced in this study, according to the law of diminishing marginal utility; then a scenic spot utility function was proposed. With the aim of maximizing the utility value, a mathematical programming model was developed. Through the process of solving the example problem, it was shown that, in a certain number of tourists, scenic managers can optimize each visitor route to upgrade the tourist experience by effectively managing the flow of visitors.

Keywords: utility; tourist flow control; scenic management; tourism modelling.

1 Introduction

Aiming at the problems in scenic like appearing overcrowded in peak tourist season, hot spots jammed and traffic jams etc. Management authorities issued a variety of responses, such as limiting the number of visitors, adding personnel to guide and maintain order. But in fact, in the face of turbulent flow of tourists, it is ineffectively just to limit the number of visitors or to guide passenger flow with experience. Therefore, to explore the flowing law of tourists in the scenic area, and to develop scientific controlling measures, become the key to solve the problem. This is not only beneficial to improving the quality of tourists' experience and travel services, but also having great significance for environmental protection and sustainable development of scenic area.

Activities of tourists in the scenic area are related to the protection and resource utilization of the scenic spot. The issue of tourist destination exactly how many people can bear have attracted the attention of ecology and tourism academic circles since very early [1]. Zhang [2] has carried out the measurement of tourist capacity on case of the scenic spot, attaining numbers such as the best capacity, the reasonable capacity, and ultimate capacity. However, in front of the tourists, we should use these numbers to limit tourists or improve management level considering the reality of tourist demand, to achieve the harmonious development of tourism and the environment. How to regulate tourists flow becomes the new problem to be solved. Some scholars have launched research. Niu [3] analysed time and space distribution of the flow of tourists in Beijing and put forward controlling countermeasures. Zhang [4] summarized factors

influencing the scenic spot by tourists and concluded they can adjust the time and space distribution of tourists by limiting amount and control tourist flow by improving the management level and other aspects. Wang [5] discussed the problem of controlling the capacity of the Tianchi lake. Feng Gang [6] used Nuorilang restaurant Jiuzhaigou scenic area as an example for empirical research and put forward a series of measures such as splitting flow of tickets entrance guard, emergency and prices and so on. Qiu [7] made Jiuzhaigou as the research object, and have carried on the exploration for a long time, providing ideas of controlling tourist flow for the scenic area. But there are a lot of studies are built on sites based on a fixed capacity. Academic circles have been debated about how much on earth the capacity of "fixed" [8]? Simon [9] concluded that eight problems existed in the conduction, including making sure that capacity are neither static nor fixed. To sum up, the present study of tourism carrying capacity account more, but the tourism carrying capacity involves less. Tourists flow regulation of the existing research is on the premise of static and fixed capacity value and neglected the tourism carrying capacity varying from person to person, and becoming different as the management change.

The following research introduce the utility concept in economics to scenic tourist regulation, set up the utility function closely related to the number of tourists; and scan the process of tourism experience in the whole tour in the perspective of utility instead of the capacity of tourist; The system integration of tourists flow control model was established, the model provides new train of thought for the peak scenic tourist flow regulation. The rest of the article knot is as follows: the second part introduces the concept of the scenic spot of utility; The

* Corresponding author - e-mail: zslu@scu.edu.cn

third part is the tourists control model and its calculation; The last is the conclusion and prospect.

2 The Problem’s Description and Definitions

2.1 UTILITY DEFINITIONS

Utility, one of the commonly used conception in Micro Economics, usually refers to a measure of consumer or enjoy leisure, to meet their own needs, desires. Visitors to meet the physical and mental pleasure purposes by the scenic tourist, scenic spot reflects its value by tourist .If the scenic area meets a person; there is a value. More people means more value. How to describe the relationship between the value of scenic spots and tourists? Under the Theory of calling the tourists’ utility as the utility of the scenic, the utility of the area will increase with the increase in visitors at first .And the increase is not a constant, but reduce with reduction of per capita occupancy resources. This change conforms to the law of diminishing marginal utility. There is a need to explain that the number of tourist also cannot be “The more, the better.” If the number of visitors exceeded a certain limit, it will take a devastating impact on environment, and a catastrophic experience to visitors. The utility will dramatically reduce then. It is out of the scope of the study to discuss on the limit of the number of visitors. So the following number of visitors is a tourist in the ‘limit the number of tourists’ within the volume. The above relationship as shown below:

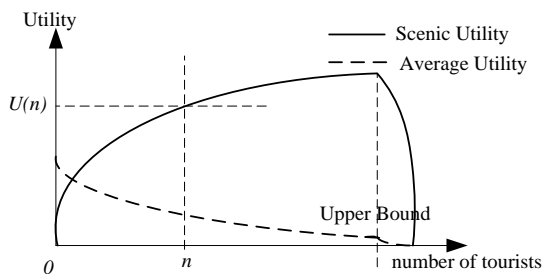


FIGURE 1 The relationship between utility and tourists’ number

Assuming the scenic value represented by the two dimensions of time and space. A scenic area is composed of a number of scenic spots; each of the spots is the space resources. Opening time is the time resource. In order to calculate the utility of the area, we should discuss it first.

Definition: The Utility of the scenic spots means the degree of an attraction being able to meet the tourists’ experience within a certain period of time and a certain number of people. It can be formulated as follow:

$$U(n) = \sum_{j=1}^n \frac{1}{j}, \tag{1}$$

where n is the number of tourists.

Suppose the visitors number is n , for each additional person lead to the total utility increases $1/(n+1)$, from $(n-1)$ to n increased $1/n$. Therefore, with the increase of the

number of people, total utility increases gradually. However, the rate of increase is reduced from $1/n$ to $1/(n+1)$, it satisfies the law of diminishing marginal utility. Suppose that there are n visitors in order to reach the spots i , leave after the stay time. The number of visitors will continue to change with time; it becomes $2n-1$ time slots. Assuming the K slot as τ_k , the number of tourists is v_k . The utility will be this:

$$U_i = \sum_{k=1}^{2n-1} \tau_k \cdot \sum_{j=1}^{v_k} \frac{1}{j} = \sum_{k=1}^{2n-1} \sum_{j=1}^{v_k} \frac{\tau_k}{j}. \tag{2}$$

Suppose the visitors’ number is m , the Eq. will be this:

$$U = \sum_{i=1}^m \sum_{k=1}^{2n-1} \sum_{j=1}^{v_{ik}} \frac{\tau_{ik}}{j}, \tag{3}$$

where τ_{ik} means the slot K at spots i . v_{ik} is the number of tourists at spots i .

2.2 THE RELATIONSHIP BETWEEN ARRIVE TIME AND THE UTILITY

Tourist scenic spot in the time of arrival time and sojourn time distribution of the number of visitors – decision. Differences in the arrival time and sojourn time lead to different scenic utility value. In order to illustrate the different effects of the utility of the arrival time, let’s make the single scenic spots as an example. Statistics shows the relationship between the time and utility.

Assume the time of every tourists to stay are the same, the opening time is 10 hours. There are 4 groups arrive in batches. The numbers of people in the groups are the same; the touring time is 2 hours. The arriving time can be the time in the opening time. In order to make the problem easy, suppose the arrival time is the whole point in 1 to 7.

Let us use four digits to express the 4 arrival time, say the number is a sequence. In this case, there are 2401 kinds of arrangement. In order to make the problem easy, there is first group of tourists arrive first hours, the next arrival time can be 0 to 2 hours after the first group. Then the number of arrangement reduces to 27.

Simulating tourists’ arriving under various time, we map the distribution of tourists arrival time and utility computing according to the Eq. (Figure 2 to Figure 4).

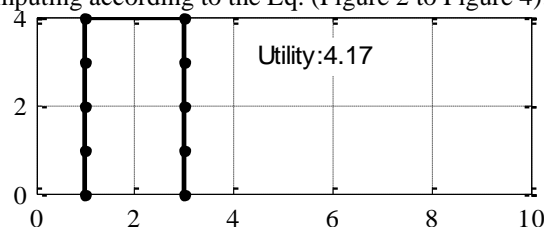


FIGURE 2 Single tourist scenic spots distribution (a, Time series: 1111)

Figure 2 depicts the 1-4 group of tourists arrived at 1:00, the number of visitors peaked 4.

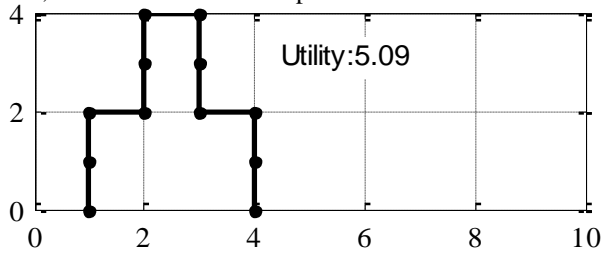


FIGURE 3 Single tourist scenic spots distribution (b, Time series: 1122)

In Figure 3, four groups of tourists arrive at the peak reached in 2 hours, 1 hour, with the first and second group of tourists leave, visitors reduced to two groups. Compare with Figure 2, although the peak value was 4, but the peak of the original duration of 1/2.

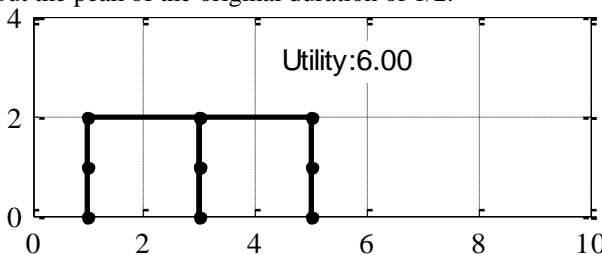


FIGURE 4 Single tourist scenic spots distribution (c, Time series: 1133)

As visitors arriving time dispersing, visitors peak decreasing. If visitors arrive at interval of two hours, then utility value reaches the maximum at this time.

The utility scenic values that the time calculated of the list are as follows.

TABLE 1 The time series and utility of tourists' arrival

Time series	Peak value	Utility	Time series	Peak value	Utility
1111	4	4.17	1124	3	6.34
1112	4	4.92	1133	2	6.00
1113	3	5.67	1134	2	6.50
1122	4	5.09
1123	3	5.84	1357	1	8.00

As can be seen from the table 1, along with changes in the timing of tourists' arrival, tourist scenic spots distributed over time changes. Case of increasing the interval, the peak reduced, the utility increased. Reduced the peak height and the peak holding time help to improve the utility, and vice versa.

2.3 THE UTILITY AND TOUR ROUTE CHOICE

Scenic area is always composed by different spots, and there is a certain degree of spatial distribution. Visitors in the scenic area choose the paths to complete the transfer between the various spots. Tourists accumulate effect in this process, as well as the utility of scenic changes.

According to the above conditions, there are two alternative paths visitors each respectively 1→2, 2→1. "1" represents the first line, "2" represents the second line. Cruise line arrangement in four groups of tourists have 2⁴=16 kinds of: 1111, 1112, ..., 2222. We take

the 1111, 1122 and 1133 from the 27 kinds timing as examples. Combination of timing and line is a total of 48 kinds of programs.

These programs were simulated, corresponding tourist scenic spots distribution (Figures 5 to 10).

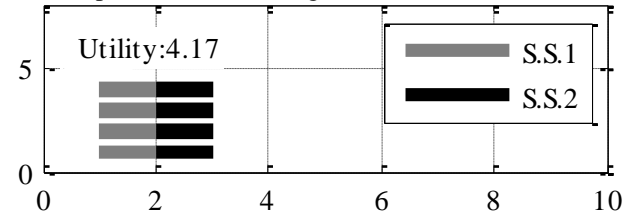


FIGURE 5 Double-Scenic spots Tourist Distribution (1a, Time series: 1111; Routes: 1111)

Both circuit arrangement and tourists reach timing shown in Figure 5 are same with what in Figure 2, the resulting utility scenic values are equal; Tourists has always maintained the highest peak 4.

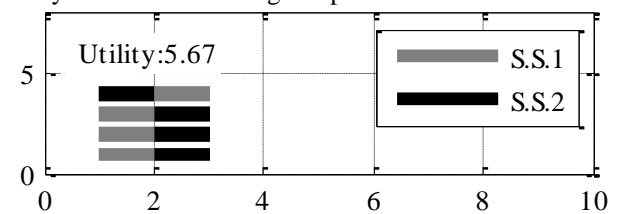


FIGURE 6 Double-Scenic spots Tourist Distribution (1b, Time series: 1111; Routes: 1112)

Figure 6 shows the front of the three groups selected Line 1, Line 2 followed by group1. As can be seen the peak of each attraction are 3, compared with the previous Routing congestion relief.

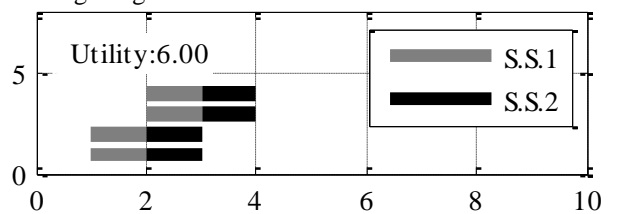


FIGURE 7 Double-Scenic spots Tourist Distribution (2a, Time series: 1122; Routes: 1111)

Figure 7 shows the distribution under the condition whose tourists arrival time meet 1122 timing, all of four groups select line1. As can be seen the peak of each attraction are 2, compared with the previous Routing congestion-relief.

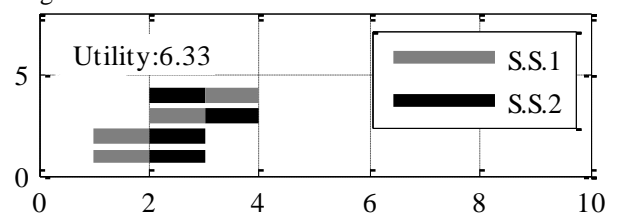


FIGURE 8 Double-Scenic spots Tourist Distribution (2b, Time series: 1122; Routes: 1112)

Chosen tour route 1112, tourist's distribution shown in figure 8 is slightly different between former, an increase of a peak at a time, the ultimate utility of a small increase.

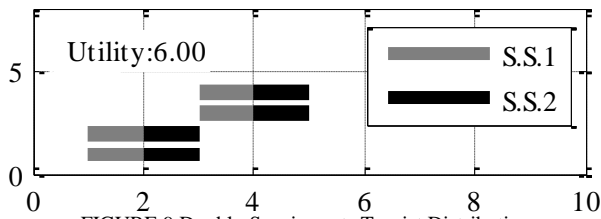


FIGURE 9 Double-Scenic spots Tourist Distribution (3a, Time series: 1133; Routes: 1111)

Under the timing of 1133, four groups tourists arrive at two batches, the arrival times of the first two groups and the latter's are shifted, no overlap, in this way, we can reach the objective of diversion, the main problem at this time focused on the inside of each batch of two group among Tourists. Obviously, you can choose a route among Tourists. Diversion program will not form a tourist's accumulation. The tourist-distribution of the peak does not exceed 2.

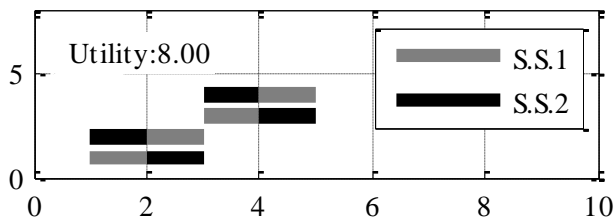


FIGURE 10 Double-Scenic spots Tourist Distribution (3b, Time series: 1133; Routes: 1212)

Figure 10 achieve the optimal combination of timing and route, and the peak of each attribution is 1. Calculate the utility under scenic route and timing combinations, and obtain the following data:

TABLE 2 The utility under part routes combine with time series

Routes	Time series		
	1111	1122	1133
1111	4.17	6.00	6.00
1112	5.67	6.33	7.00
1212	6.00	7.00	8.00
...

The table data show, choose the same timings and different tour routes, obtain different utility value; Likewise, the same path selection, because the different timing of tourists arrive, get a different tourist - distribution as well as different utility values.

Observe the above situation, if all tourists arrive simultaneously, it will cause a lot of pressure to the scenic under the first timing, and it is very important to control tourist flow by choosing tour route. Similarly, even if reached 1133, the best timing, triage has a positive meaning, by path optimization to improve the scenic utility.

There is no difficulty to infer that if the length of stay of two scenic spots are different, the transfer time between two attribution taken into account, the timing of tourists arrival time more than three kinds, then the distribution of tourists in the scenic areas will become complicated, optimizing space will be greater.

3 The model of scheduling about tourists in a scenic area based on utility

3.1 THE DESCRIPTION OF SCHEDULING ABOUT TOURISTS IN A SCENIC AREA

When tourists travel in a scenic, which consist of several scenic spots, there is a way to link. It is a kind of way to travel, such as by foot or by bus. Network is shown in figure 11. Node 1 is the point of starting and returning. There are two ways that can be chosen. One is 1-2-3-1, the other is 1-3-2-1. In order to maximize the whole utility, managers need to design the best order to go sightseeing for tourists. If there are 10 tourists and every tourist has two choices, there are 210=10*24 kinds of order combinations for sightseeing.

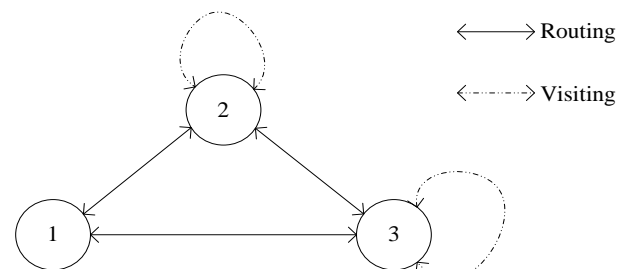


FIGURE 11 Single path network of travel mode

The fact is more complicated, the way to travel that a tourist goes sightseeing from an attraction to another attraction can also be different. (It is shown in fig.12). If there is consumption of physical strength by foot, sightseeing will be limited greatly in a relatively large scenic area. When many scenic areas provide different transports for tourists to choose, the need of sightseeing bus will appear.

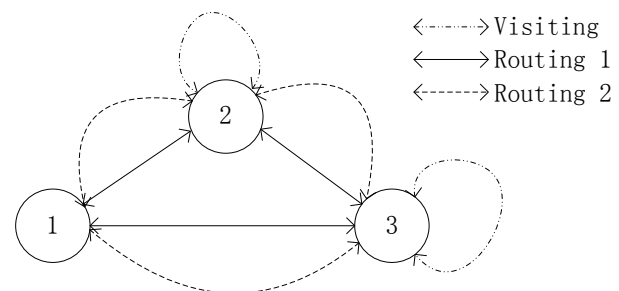


FIGURE 12 The path network Of alternative way to travel

This problem is similar to Open Shop Schedule Problem (OSSP). It is described as that n workpieces are processed on m machines. Every workpiece include m processes, every process has determined processing time. Every machine can only process a workpiece in the given time. What's more, every workpiece can only be processed by one machine. The processing sequence of workpieces on the same machine is random and the processing sequence of every workpiece is also not limited. However, the process cannot be interrupted in the processing time. The purpose of scheduling is to determine all the sequence combination of workpieces

and machines and minimize the makespan. At the same time, it has to satisfy above constraints. Therefore, the related algorithm can be used to solve the new problem.

However, the problem of scheduling sightseeing ways has its own characteristics: the resources of scenic spots belong to cumulative resources and the capacity cannot be fixed as one unit; there are differences between goal settings. The goals of OSSP usually are the whole scheduling time, the average completion time or the minimal value about other time value. However, the scheduling of sightseeing ways cannot simply pursue the short time of processing. As what above described, its goal is to maximize the utility of tourists.

So far, there are three kinds of methods to solve OSSP. They are exact algorithms (branch and bound methods, cutting plane methods), heuristic methods based on rules, metaheuristic algorithms [10, 11] and other hybrid algorithms. When $m > 3$, it belongs to NP difficult problem. Compared with the schedule problem in a shop, this problem has wider space for searching. It can use exact algorithms such as branch and bound methods to get the optimal solution at a Comparison smaller scale. But it will cost a lot of time. When it is at a large scale, it is difficult to find out the optimal solution by polynomial algorithm or pseudo-polynomial algorithm. Although the algorithms based on assignment rules is very fast, but it is usually partly optimal and the structure-dependent is strong. The heuristic algorithms based on artificial intelligence are widely used.

3.2 Tourists flow regulation mathematical programming model

To solve the above problem, establish the following planning model. Assuming a scenic area consisting of m scenic spots, the total number of visitors for the opening period is n ; S_i is the time as a starting point to scenic spots i need, but also is the time that return to the starting point of i scenic spots; O_{ij} represents the time that tourist j stay in attraction I ; scenic area pen at 0 and close at T ; Visitors j reach each scenic at time R_j ; the time to reach attraction i is a_{ij} ; and the leaving time is d_{ij} ; every tourist only visit every attraction once, but the tour order is not fixed. Scenic spots are connected by roads and trails. The ways tourists go from attraction i to the next attraction k can be walk or ride. The time need is respectively B_{ik} and F_{ik} .

When the number of scenic spots t is calculated for each time, the $m * n * T$ dimensional unit time interval index is introduced. When tourist j gets to scenic spots i at t , the value is $A_{ijt} = 1$.

When tourist j leave attraction I at the time of t , the result is $D_{ijt} = 1$. V represents the number of tourists in attraction i at t . The number of tourists at t unit is defined the number of tourists between t and $t+1$. The number of tourists at present is that the arrivals minus the number of leaving.

Attraction i is defined under time index, the calculation expression in the number of tourists at the time of t :

$$V_{it} = \sum_{\tau=1}^t \sum_{j=1}^n (A_{ij\tau} - D_{ij\tau}), \tag{4}$$

$$X_{ijk} = \begin{cases} 1, & \text{If tourist } j \text{ visit site } i, \text{ then } k \\ 0, & \text{else} \end{cases}, \tag{5}$$

$$Y_{ijk} = \begin{cases} 1, & \text{If tourist } j \text{ go to site } k \text{ from } i \\ 0, & \text{else} \end{cases}, \tag{6}$$

$$A_{ijt} = \begin{cases} 1, & \text{If tourist } j \text{ arrive site } i \text{ at time } t \\ 0, & \text{else} \end{cases}, \tag{7}$$

$$D_{ijt} = \begin{cases} 1, & \text{If tourist } j \text{ leave site } i \text{ at time } t \\ 0, & \text{else} \end{cases}, \tag{8}$$

$$\max U = \sum_{i=1}^m \sum_{k=1}^{2n-1} \sum_{j=1}^{v_{ik}} \frac{\tau_{ik}}{j}, \tag{9}$$

$$a_{ij} \geq R_j + S_i, \tag{10}$$

$$d_{ij} = a_{ij} + O_{ij}, \tag{11}$$

$$d_{ij} + S_i \leq T, \tag{12}$$

$$Z_{ijk} = (1 - Y_{ijk})F_{ik} + Y_{ijk}B_{ik}, \tag{13}$$

$$d_{kj} - O_{kj} - Z_{ijk} + M(1 - X_{ijk}) \geq d_{ij}, \tag{14}$$

$$d_{ij} - Z_{kji} - O_{ij} + MX_{ijk} \geq d_{kj}, \tag{15}$$

$$\sum_{t=1}^T A_{ijt} t = a_{ij}, \tag{16}$$

$$\sum_{t=1}^T D_{ijt} t = d_{ij}, \tag{17}$$

$$\sum_{t=1}^T A_{ijt} = 1, \tag{18}$$

$$\sum_{t=1}^T D_{ijt} = 1, \tag{19}$$

There are no special instructions in the above Eq., but the conditions are $1 \leq i \leq m; 1 \leq j \leq n; 0 \leq t \leq T$. Eq. 10 shows that the arrival time of tourist j at attraction i is greater than the time of arriving at scenic or equal to it. Eq. stands that tourists arrive at some scenic spots and finish the sightseeing and leave this attraction; Eq. 11 represents the tourist stay at the attraction for some time before

leaving. Eq. 12 shows that the sightseeing need finishing before the scenic is closed. Eq. 13 decides the way to travel from an attraction to another. Eq. 14 and 15 ($i \neq k$) define the order of visit. Eq. 16, 17, 18 and 19 are time indexes.

Supposing that tourists choose to take the single way to travel in the scenic with m scenic spots and there are interconnections between scenic spots, every tourist has $m!$ ways to travel. And n tourists have $(m!)^N$ choices. If the way to travel can be chosen (add a path between scenic spots), every tourist has $2^{(m-1)}m!$ choices. And N tourists have $(2^{(m-1)}m!)^N$ choices. Adjustable degree becomes unsolvable. Therefore, it is necessary to simplify the model. There are two ways. One is to decrease the number of indexes, the other is to decrease the base.

Firstly, the number of tasks can be reduced. The specific method is to divide tourists into n groups and arrange guide and the line for every group. This allows the total number of the order scheduling Eq. to reduce from N to I , ($n = \lceil N/G \rceil$), G is the number of every group.

According to management experience, scenic spots in the scenic are not all interconnected. What's more, the paths that tourists are like limited. Some paths can be picked out to be discussed. These paths are effective paths. So it is feasible to select k path to travel from $2^{(m-1)}m!$ paths. The k paths include the choices of transportation between scenic spots. And now, the number of solutions for the above problems reduces to k^n . It is also difficult, and it need heuristic algorithm to solve.

3.3 TOURISTS FLOW-CONTROL MODEL FOR SOLVING NUMERICAL EXAMPLE

Just as the scenic network of fig. 12 shows, Suppose there are 10 groups of tourists reach the entrance 1 in interval of five minutes, tourists start to walk or ride from 1 to 2(attraction 1) or 3(attraction 2), and return to start-point 1 after visiting 3 or 2. The time that tourists need to visit and travel is shown in the below table. The value of the diagonal in the table represents the staying time at 2 or 3. The value on the upper right of diagonal is the time of walk. The number of bottom left is the time of ride.

TABLE 3 Node time matrix

Node	1	2	3
1	0	10	20
2	2	20	15
3	4	3	30

It is easy to work out there are 16 paths to choose (shown in fig.4), 0 in the way column of the table represents walk, 1 stands ride. The first line represents that tourists choose to walk, and the order is to travel from attraction 1 to 2 and then return to 1. Every group of tourists have 16 kinds of different choices, 10 groups tourists have 16^{10} choices in all.

TABLE 4 Part of an alternative tour route list

Number	Node	Way	Node	Way	Node	Way	Node
1	1	0	2	0	3	0	1
2	1	0	2	0	3	1	1
3	1	0	2	1	3	0	1
...
16	1	1	3	1	2	1	1

Heuristics (herein abbreviated) is designed to solve the problem. The choices of 10 groups of tourists are respectively 15,7,13,2,9,13,10,2,10 and 2. Compared with the tourist group which take the fixed line (they are all choose the path 16). It means all tourists first ride to 3 and to 2, at last return to 1. Tourist scenic spots distributed as shown below. It can be found that optimized peak of tourist in both scenic spots is better than the compared program.

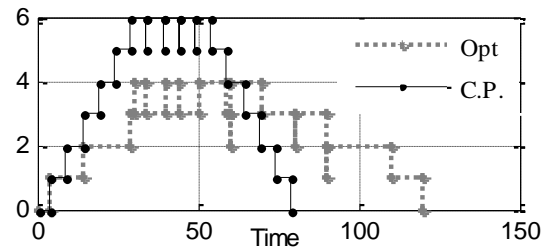


FIGURE 13 Comparison of distribution of tourists on scenic spot2

When contrasting the mean number of tourists and variance, they are also improved significantly. The utility of goals optimization reached 17.429 while the utility of the contrasted group which takes the fixed line is only 13.225. It shows that the above model is effective in tourists controlling.

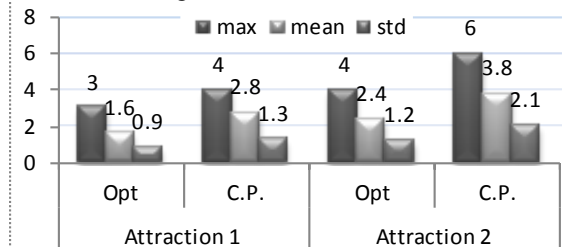


FIGURE 14 Comparison between optimized program and control program indicators

Although the model just change the path that tourists take, in fact, it adjusts the timing of tourists arrive scenic spots. And the purpose achieved by staggering the timing of changes. If the control measures are taken in the arrival rate, the distribution of visitors could be further improved. In fact the use of the vehicle can be understood as increasing the path selection, and under the information technology' support, dynamic scheduling control will make more space.

4. The conclusion and prospect

When scenic spot management authorities face turbulent flow of tourists in the tourist season, not simply because of the bearing capacity of the scenic spot turn them away. To solve the contradiction need strengthen the

environmental protection consciousness of tourists, also call for management innovation.

It is a new attempt to introduce the concept of economic utility into a scenic tourist flow controlling. It is simple and clear to set up the utility function following the law of diminishing marginal utility. The above analysis shows potential in regulating peak visitors with the model of maximizing the scenic spot of the utility. The factors influencing utility in the model include arrival time, sojourn time. In practice, managers need to focus on these two kinds of time. Although it is inconvenience to control tourists arrival time in the scenic area, but can regulate the time-space distribution of tourist flow in tourist scenic spot. Measures include offering visitors the selection of travelling way and arranging touring line. It will be of reference value to improve tourist experience. Because the scenic areas (spots) capacity itself is an adjustable object, it were deliberately ignored in the model. If scenic spot have bottleneck in the peak, the model can be applied to the scenic spots alone.

The paper assumes that the increasing or decreasing in the number of tourists in scenic spot makes the same contribution to the utility. In fact, the differences between these need researching. As to the problem of large-scale tourist flow regulation, it needs to explore more efficient algorithm.

Acknowledgments

This research was supported by the Major International Joint Research Program of the National Natural Science Foundation of China (Grant No. 71020107027), Humanities and Social Sciences project of The Ministry

of education of China (Grant no.12YJC630023), Social sciences program of Chengdu (Grant no. SKNHR13-07).

References

- [1] Wagar J A 1964 The carrying capacity of wildlands for recreation *Forest Science Monographs* 7 1-23
- [2] Zhang Xiaoping, Zhu Zhongfu 2007 Jiuzhaigou Tourism Environmental Capacity Study *Tourism Tribune* 22(9) 50-57 (in Chinese)
- [3] Niu Yafei, Xie Libo, Liu Chunfeng 2005 Beijing tourist flow characteristics of spatial and temporal distribution and control countermeasures *Geographical Research* 24(2) 283-91 (in Chinese)
- [4] Zhang Jianhua, Jian-Hui Yu 2006 Tourist attractions tourists environmental impact analysis and its regulation *Beijing Second Foreign Language Institute* 1 95 – 98 (in Chinese)
- [5] Wang Hui, Yang Zhaoping 2008 Xinjiang Tianchi Regulation of Tourism Environmental Capacity *Arid Land Resources and Environment* 22(9) 193-96 (in Chinese)
- [6] Feng Gang, Ren Peiyu, Xiaowei Yang, et.al. 2009 *Research on Ecological Protection scenic RFID-based mode – Jiuzhaigou Nuorilang restaurant dining diversion for example* Southwest University for Nationalities, Humanities and Social Sciences Edition 11 151-54 (in Chinese)
- [7] Qiu YanQing, Ge Peng, Liu Zhusheng, et al. 2011 Research shunt Tourists and space-based navigation management control of complex systems *Soft Science* 25(9) 54-57 (in Chinese)
- [8] Lindberg K, McCool S, Stankey G 1997 Rethinking carrying capacity. Research Notes and Reports *Annals of Tourism Research* 24(2) 461–65
- [9] Simon F J G, Narangajavana Y, Marques D P 2004 Carrying capacity in the tourism industry: a case study of Hengistbury Head *Tourism Management* 25 275-83
- [10] Liaw C F 1998 An iterative improvement approach for the nonpreemptive open shop scheduling problem *European Journal of Operational Research* 111 509-17
- [11] Gao Liang, Gao HaiBin, Zhou Chi 2006 Based on open-shop scheduling particle swarm optimization *Journal of Mechanical Engineering* 42(2) 129-34

Author



Liu Zhusheng, March 16, 1969, Chongqing China

University studies: Ph.D., student of Sichuan University.

Scientific interest: computer simulation modelling, planning and scheduling optimization in manufacturing and services.

Publications: 10 publications.

Experience: Computer Integrated Manufacturing System; Applied Research of RFID Technology in the Nature Reserves & Earthquake Ruins and Sites based on Spatiotemporal Separation Management Pattern, which are projects of National High Technology R&D of China (863 Program).

The agent – based warning modelling of internal quality risks in supply chain for manufacturer

Liu Caihong^{1, 2*}, Xiong Wei¹

¹ School of Management, Zhejiang University, Hangzhou 310000, China

² School of Business, Jiaying University, Jiaying Zhejiang 314001, China

Received 1 January 2014, www.tsi.lv

Abstract

With the dynamic of market and deepening competing, enterprises will to be faced with much uncertain quality risks in supply chain. At present, the quality risk has been seen as one of the most important risks of supply chain, which is the most difficult to prevent and manage. To do it, the entities-attributes model based on the business of manufactures was used as the reference data sources of building evaluation index system for internal quality risks of supply chain. Because the early warning of risks manually has the limits of information insensitivity and risks identification slowly, this is apt to cause the delay of risks precautions. Therefore, the intelligent Agent based modelling method will be applied to construct the four early warning situations based warning model according to the internal risk of supply and demand for manufacturers. In addition, the evaluation algorithm, the rules of early warning and the running process for this Agent were focused in the study. This study will play a reference role to the analysis and management on the quality risks of supply chain.

Keywords: Quality risks, Supply chain (abbr. SC), Agent, Manufacturing enterprise, Early-warning modelling

1 Introduction

Because of the anfractuositities of structure, instability of environment, managers pursuing lean supply chain and neglecting risk management, and so on, supply chain is having been more and more subject to affected by all kinds of risks [1].

Presently, the supply chain is the most favourable management model for enterprise. Especially, the manufacturers businesses are mostly operating on the supply chain. However, with the widely using of supply chain in the management of manufacturing enterprises, enterprises mainly focus on their core business, and outsource their non-core business to other companies, it can help enterprises to do fine, but due to the lack of supervision on the other outsourcing and the complex of market, we will encounter many problems about the quality of products [2]. So is it that we will to be facing with more quality risks of supply chain for manufacturers.

Furthermore, we are being the active period of reform and opening up in China, which makes manufacturers in China, faces more uncertainty. As a result, some unexpected events from uncertainty become more frequently, such as, the “melamine milk” incident in 2008, the capital chain rupture in enterprise in 2011 made Wenzhou in Zhejiang province encountering great economic crisis, the lean meat powder event in China caused huge losses to the pork market, etc. These are the typical quality risk event in supply chain. In addition,

China is becoming the world's manufacturing plants, to improve the Chinese' international competition in products, markets and technology, some new challenges for SC are increasing. The study on it has become a new hot in the field of supply chain management in China.

Under the environment of Globalization, many risk in SC for manufacturer come from these relation interrupting of supply and demand. Such as, the third party intervention, strike, natural disasters, human error [3], the habits change of customer, technical failure, financial difficulties, and etc. these problem can be seen as resulting from the external cooperation based on resources of supply and demand and internal management. Namely, enterprise is challenged by the supply chain quality.

Although it is all but impossible to predict a specific risk case. However, it is feasible to permit us to predict the risk patterns or trends in the evolution of sprout from the perspective of macro management. In view of the quality risk in SC for manufacturers, if no natural disasters and political change, should to be formed by the internal quality of manufacturer and the quality of cooperation with others in marking. It is to say that this risk can arise from the internal of enterprise and the external interface business. From the point of SC system, the external interface businesses to manufacturers are based on the supply and demand cooperation between manufacturers to other nodes in its SC and nodes in other SC. In view of this, on the premise of no big natural disasters and political reform, we will analyse the quality

* Corresponding author - e-mail: liuch0001@163.com

risk factors from the internal operation of enterprises and external interactions for all kinds of supply and demand, furthermore, the Agent based model of predicting quality risk will be provided to recognize the possible macro risks related to the quality of enterprise.

2 Literature review

With the integration of various management concepts, risk management theories begin to be applied in quality management, which makes more scholars to research the quality risk in various fields. Furthermore, in the competition environment of the high and dynamical requirements of customers on product quality and service level, the quality risks in SC has become one of the bottlenecks of enterprise operation. Although, the study on it has become hot, but superficial.

Levy (1998) [4] found in the environment of SC, the quality had a paradigm shift. Quality problems are not only in internal of enterprise, but also in these aspects of customers, supplier, and collaboration to ensure the quality of the products. Chu-Hua Kuel (2003) [5] and etc., by comparing and analysing the supply chain quality management, total quality management and comprehensive quality assurance, provided the key factors of risks in SC are composed of information, relation, and innovation are analysed, that information, and the relationship between the reform and the three is a key factor to guarantee the quality of products, reduce the quality risk in supply chain. Svensson G. (2000) [6] discussed the supply risk in production logistics, and proposed a conceptual framework of supply chain vulnerability analysis. As a famous expert in supply chain management, Christopher S.T (2006) [7] studied the supply chain quality risk management from the four aspects of product, supply, demand and information. Batson & McGough (2007) [8] build quality model of supply chain for quality prediction and quality improvement by the ideas of quality engineering. Dahiya et al (2009) [9] studied a HACCP (The Hazard Analysis and Critical Control Points) system, It aimed at preventing and reduction these known risks in physical, chemical and biological hazards by identification, assessment and safety control.

In the last ten years, in China, the study on supply chain quality risk can to get attention. The typical research mainly presents the following. Zhou Chao-yun, Lu Zhi-qiang (2011) [10] studied the supply chain operation strategy with product quality risk and transport service quality risk under the condition of information asymmetry, and used three-stage dynamic game theory to study how the members of the supply chain constitute reasonable strategies to make sure that they can gain maximum profits themselves while facing the asymmetrical quality risk. Ye Han-yan [11] studies some theories about supply chain risk management and presented that the disruption risk and forecast risk are two most important risk of supply chain. Gu Li-gang, Gao

Tao (2009) [12] studied the coupling mechanism of supply chain quality risk and its characteristic, and put forward a quality risk disruption management which faces to the final products from core conception level and operational level respectively. Qian Ying (2007) [13] in her paper discuss the rules of quality flow in supply chain and how supply chain members collaborate and coordinate with each other to achieve the synergetic operation of supply chain quality system. Yan Zhong-e (2013) [14] from the view of customer satisfaction deviations, studied the formation mechanism of supply chain quality risk and provide the quantitative analysis model of quality risk transfer in SC based on quality function.

The above study observed that the analysis of risk is not much unified, the combined SC risk and enterprise risk is lack. As a result, we obtain the factors of risks relatively poor in effectiveness. Moreover, due to the lack of risk prediction methods and techniques, the research in this area is still in the basic theory, such as the connotation, characteristics, risk evaluating, or a certain level of quality problems (such as risk analysis of supplier), and so on. There is one more point, the study of SC quality risk is still at an initial term, there is not a accepted concept by most people, so is it that the SC risk and quality risk of SC is easy to be confused.

To help more enterprises developing smoothly, it is very necessary to study risks management, in which early warning of risks is essential to enhance the capability of recognition risks. Therefore, here, the study on intelligent early warning assisting the operation of SC in manufacturer successfully will be focused on.

3 The Core Quality Risk Fields of SC for Manufacturer

3.1 CONNOTATION OF QUALITY RISK IN SC

Quality risk refers to the occurrence possibility and consequences of negative quality events in a certain time. It is that can lead to a certain quality loss or other adverse consequences. Under the supply chain, the quality risk of enterprise is sum of the risks in all aspects of supply chain [15-16].

Most experts believe that the quality risk in SC is more difficult to be prevented and be managed. Compared with other risks in SC, the quality risks have some unique characteristics [17]:

- 1) Transitivity: Any small perturbations on supply chain may take "the Butterfly Effect" to itself. This process is non-linear with all kinds of unexpected happens that it can lead various quality risks to us.
- 2) Time lag: Quality risk derived from various kinds of quality problems. When time was is enough, risks latent in these questions are to be burst out. Therefore, quality risk of supply chain need undergo a complex evolutionary process, but not

timely.

- 3) Complexity: The quality risk of the supply chain is a new dimension of coordination management in supply chain. Compared to the three dimension of the price, order quantity and delivery time, quality dimension has greater complexity in the supply chain coordination.
- 4) Interaction: on the role of transitivity, there are also interactions between the quality of supply chain risk and other risks in supply chain.
- 5) Easily happening: In the complex and dynamic business environment, supply chain quality risk is prone to happening. And so on.

These characteristics are regarded as "black box of quality" for further increasing the difficulty of quality risk management.

3.2 ANALYSIS ON CORE QUALITY RISK AREAS IN SC

Supply chain is formed by the activities of supply and demand, including of raw materials procurement, production, processing, sales and other business behaviours. It is every step has itself procedures and has interaction with other steps. In the open complex and dynamic environment, SC is actually a resources network of supply and demand [18]. Each node in the SC is the host of supply or demand. Resource flows between nodes in fact mapped out a logical relationship between the nodes, it is that the forming of the supply chain structure [19]. If the strategy of node in SC is fully open cooperation, SC should to be a dynamic and opening network, which is more in line with the actual market environment. So, here, the study is based on this supply and demand network to analyse the risks in SC for manufacture. As a manufacturing enterprise, from the internal, its quality is composed of the enterprise itself and the supply and demand relationship (briefly, the cooperation on all kinds of resources in SC). At these premise, we propose the following analysis.

Firstly, as for the quality in the manufacturers, it is depended on the ability of quality management. Such ability is mainly manifested in: the R & D of enterprise, the production, the average knowledge level of employees, management skills, service, equipment capacity, quality of supply chain information communication, the docking of technology platform and so on. Such as, manufacturer, in the development and design of product, is no clearly grasping the supplier's ability of meeting the requirements of new product. As a result of the blind cooperation, risks in product might be brought about for the inferior raw materials. So, the quality risk in SC, which caused by the defects of quality ability and the defect of information communicating on quality, will reduce the performance of cooperation.

Secondly, In terms of the cooperation risk in enterprise mainly arising from the cooperation contract of collaborators, the credit is the key of cooperation successfully. Zhang Hai-feng [20] studied the credit risk sources of enterprise, and pointed out if the co-operators are unwilling or unable to fulfil the stipulated quality in contract, the credit risk of SC can to be come from the corrupting influence of moral on each other, including three kinds of situations: one is that the contract party has the ability of supply or demand, but has no intention to perform the contract; the other is no ability but is willing to fulfil its obligations; the third is that partners have neither the capacity to carry out the contract, but also unwilling to carry out contract.

On the above analysis, the early-warning issue areas are suggested as U , $U = \{U_1, U_2, U_3, U_4\}$, U_1 and U_2 presents respectively the warning conditions. In which, $U_1 = f (f' (\text{index}_1) \times f' (\text{the weight of index}_2), \dots, f' (\text{index}_n) \times f' (\text{the weight of index}_n))$, With the same method, the values of U_2, U_3, U_4 will be worked out. It shows that the key of the early-warning lays in the construction of the indexes system and the obtaining of index values.

According to the traditional setting method of index, such as the risk classification based on enterprise business process and the sector process, etc., we will obtain a complex index system. It was not conducive to the core index, but also not conducive to the judgment of early warning. Considering the quality risk of manufacturer in SC mainly is focused on the perspective on the ability of quality management of enterprise and on the ability of supply or demand of collaborator. That is to say, these aspects are the two core source ways of core quality risks in SC for manufacturer. So, the evaluation indexes can be transformed from some of the behaviours properties of the two kinds of risk source. To do so, the model of E-R as an effective method is applied in the any sis of attributes. Therefore, the source of risk index will rely on the E-R model of the supply and demand between enterprise and its co-operators.

4 Setting Up the Evaluating System of Quality Risk in SC for Manufacturer

4.1 THE MODEL OF ENTITY-RELATIONSHIP (E-R)

The E-R (Entity-Relationship) model is also named as the conceptual data model. According to the core field of risks, the core behaviour of a manufacturing mainly is included in these aspects of service providers, production and supply and demand of resources based on strategic partners. Based on this, the entity-property in the conceptual data level is shown as the following figure 1.

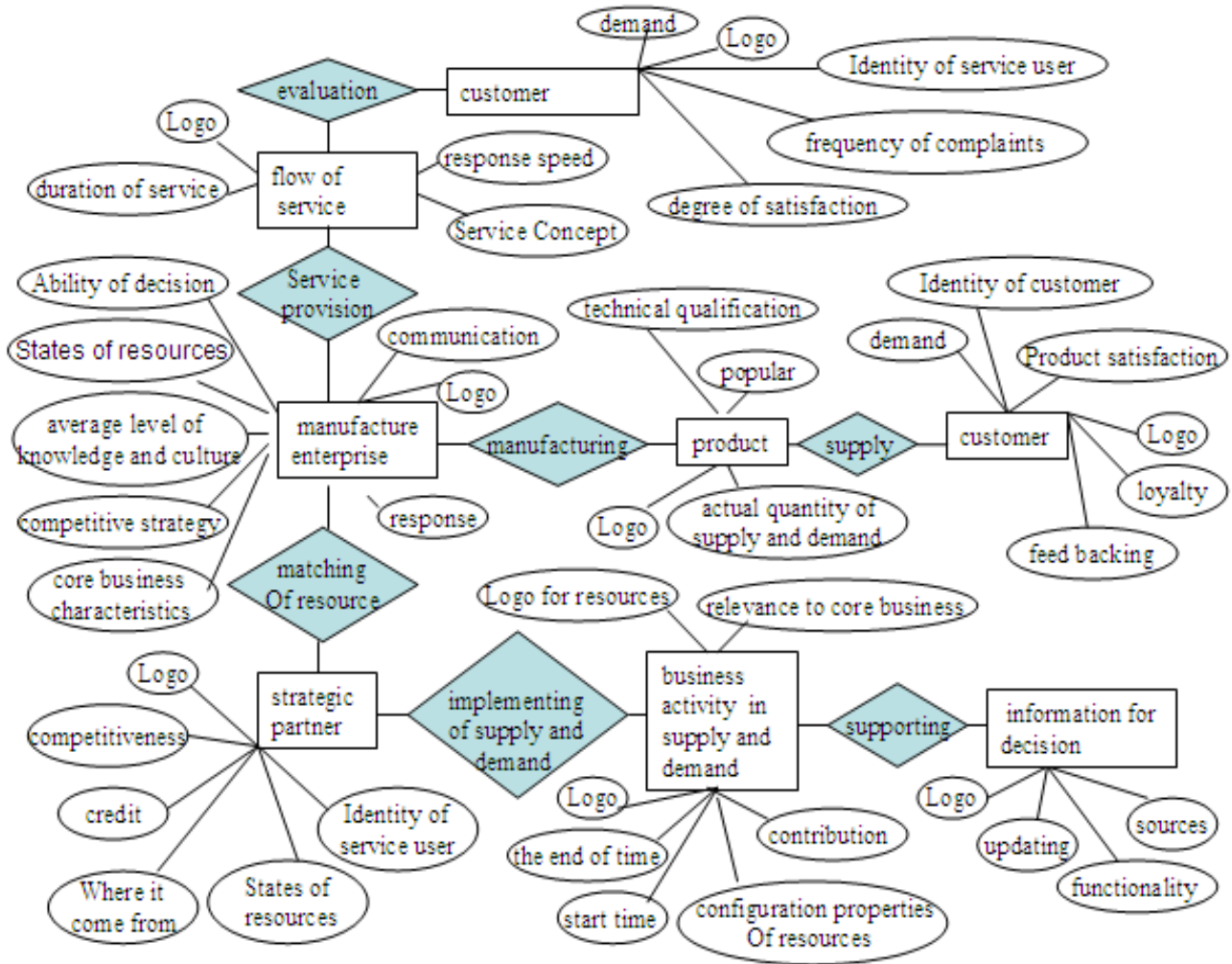


FIGURE 1 The E-R model of manufacturer

In figure1, the rectangle marks the entity; the ellipse represents the attributes of entities; the diamond indicates the links between entities. Here, we think evaluation index as another style of entity' attribute. According to the relationships attributes, the early-warning indexes can be set, which not only simplifies the complex issues of the index origin, but also endows the following early-warning index system with certain objective evidence based on the conceptual model of database.

4.2. TRANSFORMING THE E-R MODEL INTO EARLY-WARNING INDEX SYSTEM

Thinking the Entity-Attributes reflect in the quality of entities and the SC quality for manufacturing, and the conceptual entities in figure1 are the source of database, it can provide some meta-level data parameters for the risk management of enterprise. As for the warning, the early warning relies in the monitoring that is indispensable to the evaluate indexes [21], and the evaluate indexes are parameters from attributives of entity behaviours. So, the index system of risks early warning is constructed as shown in figure2.

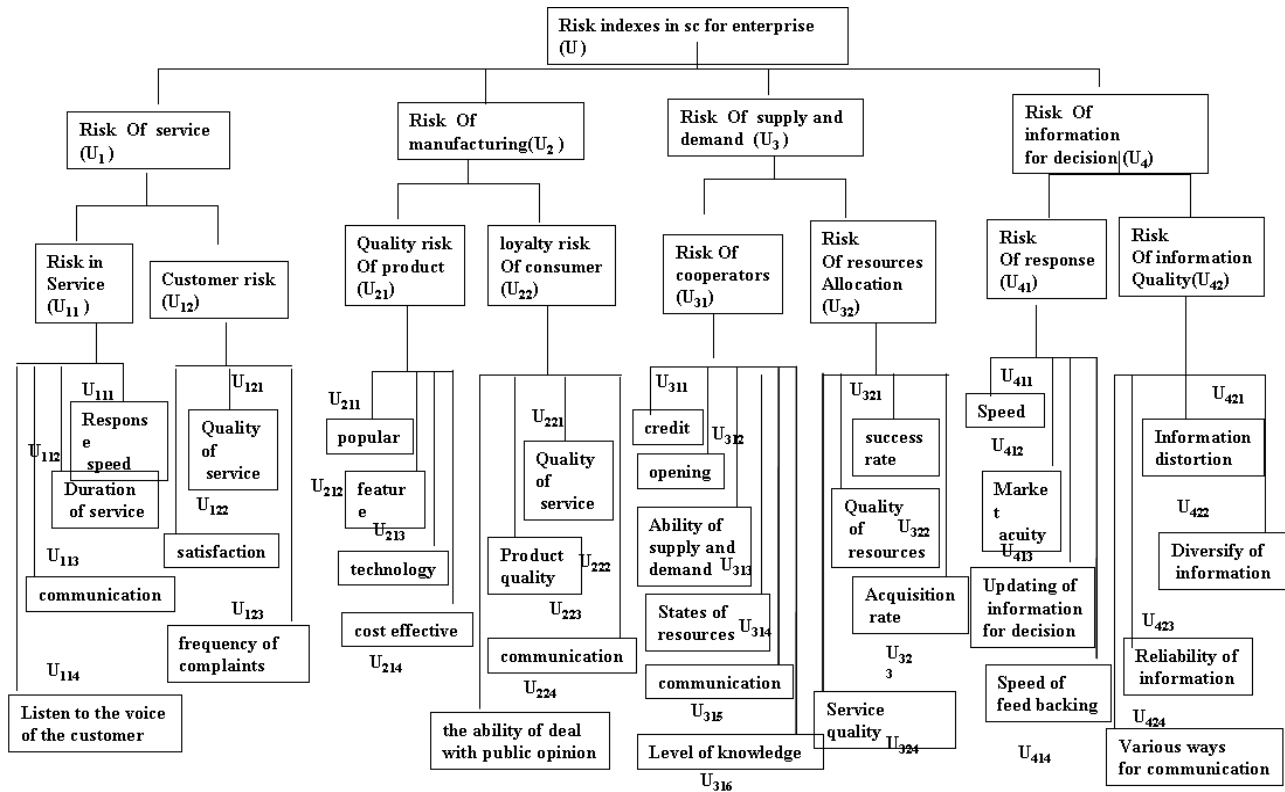


FIGURE 2 The index system of risks early-warning

From figure2, as can be seen that risks in SC for manufacturer lies in these aspects: The cooperation risk (partnership, cooperation content risk, etc.), configuration risk of resources (logistics risk, etc.), after-sales service risk (public opinion risk, customer loss risk, etc.), and the risk of quality management (correct and timely decision-making, information integration ability, knowledge and cultural lacking, etc.).

5 The Agent-based Modelling of Quality Risk Early Warning for SC of Manufacturer

5.1 THIS STRUCTURE OF EARLY-WARNING AGENT

Agent is an intelligent system with the characteristics of autonomy, reactivity, initiative and interactivity, which can solve some comparatively complex problems. The intelligent decision builds on the structure of Agent [22], which usually has the cognitive model, sensor model and the mixed model. In terms of this transformation, the early warning is a serial complex behaviour process involving information processing, information exchange and reasoning. Therefore, the E-W Agent will be constructed as mixed model shown as figure3, which is usually composed of three parts of perception, cognition and output [23-24].

The main task of the Agent structure is to solve what modules they are composed of, how the modules exchange information, how the perceived information influence its behaviours and the inside situation as well as how these modules are combined to form an organic

whole through certain software and hardware so as to realize genuinely the subject. In figure3, the function of the component parts is listed as follows.

(1) Perception. Some information produced in the supply-demand activity are obtained through the sensors of all kinds of supply-demand activity, which include the names of the supply-demand flow, the ways of supply-demand (participant nodes), the fluctuating information of the supply-demand and the information of the profit and loss in the supply-demand. All kinds of the data obtained through the sensor will offer the lower data support to the further settlement of the information.

(2) Information processing. The input information by outside environment as primarily information is sorted by the thought of Figure2 to find out the risks inducement from the mixed information. Because the inducement information is generally obscurely described, it is difficult to estimate the harmfulness of the risks in quantification and to classify the risks in the qualitative aspects. Because the inducement is indefinite and the expression is incompletely identical, it is impossible to mark the all the inducement information separately with the quantitative method. Then the inducement information can only be classified and analysed on the consideration of the classification of resource, so the Agent can only predicate the SDN core risks in a rough way. By this way, the value of risks is analysed to the early-warning value on the basis of the strategy of estimating risks and then the corresponding early-warning information will be output. In this sense, only the origin path and scope can be predicated.

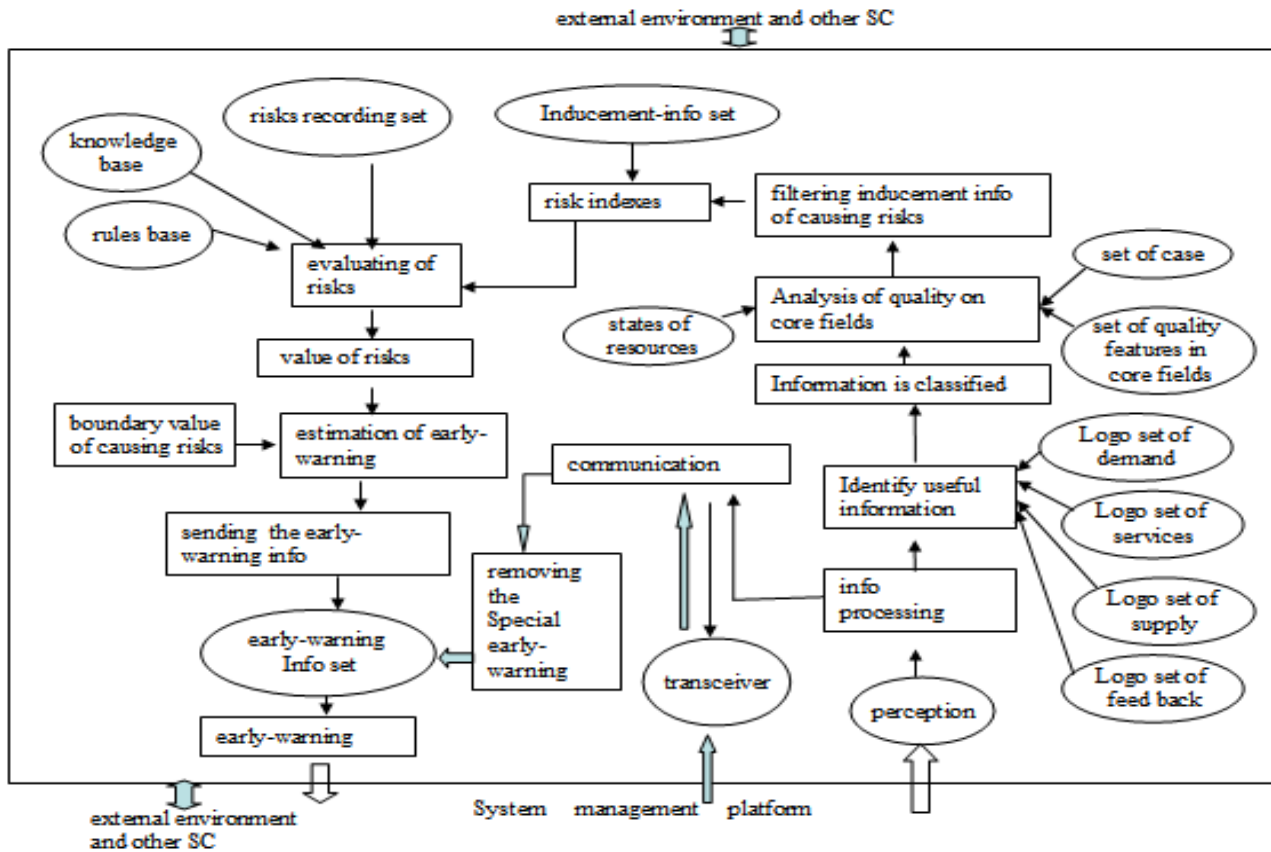


FIGURE 3 The structure of this early-warning Agent

(3) Base of knowledge and culture is compose of these knowledge of quality risk in SC, evaluating risk, the core risk fields, classification knowledge, supply and demand knowledge, management knowledge, culture of enterprise, and so on.

(4) The Class Define of Data for this Agent
According to figure3, the data class of the Agent is defined as follows:

```

Class Agent
{ private:
    Strategy of cooperation A[2]; /* full-opening or semi-opening*/
    quality feature qf[4][10]; /* the set of quality feature for service, supply, demand, and info */
    rule-base rb; /* the rule base of reasoning or the alerting threshold ascertained*/
    knowledge-base kb; /* knowledge base */
    index-kb ikb; /* the knowledge base of evaluating index systems */
    case-base pc; /*the case base of early-warning*/
public:
    database db; /*database for system*/
    business bs; /*the base of business*/
    core-task ct; /*the core business of enterprise*/
    resources-base rb; /*the sharing set of resources*/
    strategic-co-objects sco; /* the library of strategic cooperation objects */
    resources state s [2][6]; /* the quality states of all resources for manufacturer and collaborator*/
    index-base ib; /*the base of evaluating indexes*/
    logo-set Ls[4] /*logo set of four core field*/

    Communicator M-communication // the class of communicator
    Void infoobtain(); //the function for filtering the information of indexes
    float second-index-important(name,value); //the function of evaluating weights of indexes
    float jingqingjisuan(); //the function for calculating warning
    Void jingdupanduan(); //the function of reasoning alerting threshold
    Void SearchDataSet(); // the function for searching database of system
    int states of quality(); //analysis of quality in core fields
    Void SearchCaseSet(); //the function for searching case-set
    ... //other defining of extending
}
    
```

(5) Databases constructing

Supposing a manufacturer enterprise as “M-E”, according to practical situation, databases to be used are described informally as follows:

Logo bases::=<base mark, classification of demand, supply classification, service classification, information classification, information source direction, expanding base>.

All resources bases in “M-E”::=<base mark, classification of supply-demand flows, identity of resource possessor, name of resource possessor, quality of resource (worse: 0; general: 1; better: 2; very good: 3), quantity of resource (surplus:1 or loss:0), the current utility state (possible: 1, impossible: 0) of resource, expanding base>.

information database::=<base mark, key words, information integrity (good: 3, general: 2, poor: 1, very poor: 0), information authenticity (good: 3, general: 2, poor: 1), contents, information-source, contribution for decision, expanding base>.

strategic co-partners base::=<base mark, name of co-partner, identity of co-partner (as to say role mark), resources base of co-partner, credit of co-partner, competitiveness of co-partner in peers, cooperative strategy of co-partner, cooperative state of co-partner (external, strategy, under the relation), expanding base>.

information base of roles being required::=<base mark, role mark, the main supply-demand flows marks in roles, the pattern sets of the main supply-demand flows in roles, expanding base>. Different co-partners should to act different roles in performing certain supply-demand task such as role of material, role of sale, role of human resource, role of funds, role of technology and so on. These roles separately correspond to the flowing of resources on material, products, human resource, funds, technology etc.

core business base of “M-E”::=<base mark, business mark, supply-demand flow mark, possible roles describing, its main function, sets of roles required in core business (such as production, manufacture, sale, scientific research, talent cultivation, purchasing, information collection, capital supplying, management etc.), the actual shapes of certain supply-demand flow generating from a certain role (such as, the resource flow of cultivated talents can be classed practically by top manager, fitter technician, miller technician and so on), expanding base>.

informing base of new supply-demand business::=<base mark, business mark, mark of role under the business, mark of main supply-demand flows, identity of enterprise, co-partner sets, the predicted transaction value, the predicted payment cost, time for starting, expanding base>.

base about supply-demand relations::=<base mark, classification of supply-demand relations, the horizontal

supply-demand relations base indicating the cooperation between competitors, the vertical supply-demand relations base showing the compatibility of different roles to a systematically task, the crossing relations base displaying the complexity of this supply-demand relations, the hub node with some different properties neighbour nodes, times of terminating of supply-demand relation positively or passively in activities, expanding base>.

base of horizontal supply-demand relations::=<base mark, horizon function mark, the affiliated business mark, the activated state of relations (active or dormancy), sets of participants, mark of this main supply-demand flow, cooperative strategy being adopted (alliance with finite scope or “all comers are guests”), the invalidation relations, times of terminating of these relations positively or passively in activities, expanding base>.

base of vertical supply-demand relations::=<base mark, vertical function mark, the affiliated business mark, the activated state of relations (active or dormancy), sets of participants, mark of this main supply-demand flow, cooperative strategy being adopted (alliance with finite scope or “all comers are guests”), the invalidation relations, times of terminating of these relations positively or passively in activities, expanding base>.

base of early-warning cases::=<base mark, inferential rules, warning condition, warning level, key words of contents about risks, expanding base>.

base of boundary value of risk: =<base mark, boundary rules, classification of boundary, boundary value, boundary warning, expanding base>. In figure1, the rectangle marks the entity; the ellipse represents the attributes of entities; the diamond indicates the links between entities. Here, we think evaluation index as another style of entity’ attribute.

(6) Evaluation Algorithm of Index

Suggest the weight of the second-level indexes is considered, and the weight values are respectively taken as K_{11} , K_{12} , K_{21} , K_{22} , K_{31} , K_{32} , K_{41} , K_{42} . The evaluation degrees of the weight are set like this: very important (3), comparatively important (2), general (1), unimportant (0). The experts give the specific values after the comprehensive evaluation. Except that the index weight is evaluated subjectively, the evaluation is made by extracting the objective data of the indexes with consideration of the above listed database in order to evaluate the warning condition objectively. The descriptions of all the warning conditions and index evaluation are as follows.

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

TABLE 1 U_1 and its index evaluation

U_1	$U_1 = K_{11} \times U_{11} + K_{12} \times U_{12}$
U_{11}	$U_{11} = (\log_{(U_{112} \times U_{113} \times U_{114})}^{U_{111}}) \div DV$. In which, the character DV is supposed as the desired value of U_{11} . If $U_{11} > 0$, then the risk boundary value of $U_{11} = 1$; otherwise, the boundary value of $U_{11} = 0$.
U_{12}	For the U_{122} and U_{121} respectively has an inverse relationship with U_{12} , and U_{123} is positive proportional to U_{12} . Then, $U_{12} = (U_{123})^{U_{122} \times U_{121}}$.

It can be concluded from the table1 that $\max(U_1) = K_{11} \times 1 + K_{12} \times 1$, $\min(U_1) = 0$.

TABLE 2 U_2 and its index evaluation

U_2	$U_2 = K_{21} \times U_{21} + K_{22} \times U_{22}$
U_{21}	Supposing Q as the desired value of U_{11} , then $U_{21} = Q \div (U_{211} \times U_{212} \times U_{213} \times U_{214})$
U_{22}	For the range of each sub-index value in U_{22} is set four stage (very high: 3, high: 2, general: 1, poor: 0). Then, $U_{22} = U_{223} \div (U_{211} + U_{212} + U_{213} + U_{214}) \times 10$. If $U_{22} > 2$, then the risk boundary value of $U_{22} = 0$; if $1 = U_{22} < 2$, then the risk boundary value of $U_{22} = 1$; otherwise, the boundary value of $U_{22} = 2$.
U_{23}	Suppose three kinds of state information in the active supply-demand for the strategic objects: unused, using, and used. There are Q objects recorded in the database with Q' objects unused, then $U_{23} = Q' / Q$, which shows the use utility of the nodes in the supply-demand relation structure. $0 \leq Q' \leq Q$, so $0 \leq U_{23} \leq 1$.
U_{24}	Suppose H_k as the number of the role relations of the supply-demand with abnormal termination in certain supply-demand behaviour, H_k' as the number of all the role relations in the supply-demand activity and h supply-demand businesses running in the current enterprise, then $U_{24} = (\sum_{k=1}^{k=h} H_k / H_k') / h$, which represents the abnormal supply-demand phenomena for the current average appearance. $0 \leq U_{24} < 1$.

It can be known from the description on the index valuing of the warning condition U_2 in the table2 that $0 \leq U_2 < (4 \times K_{21} + K_{22} + K_{23} + K_{24})$.

TABLE 3 U_3 and its index evaluation

U_3	$U_3 = K_{31} \times U_{31} + K_{32} \times U_{32}$
U_{31}	Suppose H_k as the number of the relation quality of the supply-demand with abnormal termination in certain supply-demand behaviour, H_k' as the number of all the role relations in the supply-demand activity and h supply-demand businesses running in the current enterprise, then $U_{31} = (\sum_{k=1}^{k=h} H_k / H_k') / h$, which represents the abnormal supply-demand phenomena for the current average appearance. $0 \leq U_{31} < 1$, in which, $H_k = 1 \div (U_{311} + U_{312} + U_{313} + U_{314} + U_{315} + U_{316})$
U_{32}	V_{yT} and V_{yt} represent the quality states of the high-quality resources V_y at the time of T and $t (T > t)$. If there are m high-quality resources in the current enterprises and $(\sum_{y=1}^{y=m} (V_{yT} - V_{yt}) / \sum_{y=1}^{y=m} V_{yt}) > 0$, then it means some high-quality resources loss, $U_{31} = 1$; otherwise, $U_{31} = 0$, which means no risks. And, $V_y = (U_{324} \times U_{322})^{(U_{321} \times U_{323})}$.

TABLE 4 U_4 and its index evaluation

U_4	$U_4 = K_{41} \times U_{41} + K_{42} \times U_{42}$
U_{41}	Suppose three kinds of state information in the active supply-demand for the strategic objects: unused, using, and used. There are Q objects recorded in the database with Q' objects unused, then $U_{41} = 1 / (U_{411} + U_{412} + U_{413} + U_{414}) \times Q' / Q$, which shows the use utility of the nodes in the supply-demand relation structure. $0 \leq Q' \leq Q$, so, $-1 \leq U_{41} \leq 1$.
U_{42}	If the number of info-flows is g' and the number of obtaining information is g , then $U_{422} = g' / g$. According the role of sub-index, then, $U_{42} = \sqrt[4]{(U_{421} + U_{423} + U_{424})}$.

The evaluation is made by extracting the objective data of the indexes with consideration of the above listed database in order to evaluate the warning condition objectively.

6 Conclusion

The study on quality risks in SC is hot, but difficult. The main contribution of this study is to elaborate the predicating thought of core quality risk in SC and the structure of early warning Agent, put forward the attribute definition, data class definition of Agent and database for early warning and evaluation algorithm of index. This study will provide certain theory and practice preference to following up development of this intelligent early-warning system. However, it needs further researching in the following research.

Acknowledgments

The authors wish to thank Xu Fuyuan, Lu Jianshe. This work is supported by the natural science foundation of Zhejiang province in China (Grant No.LY13G010005), supported by the major program of the National Social Science Foundation (Grant No.12 &ZD206) and supported by the postdoctoral foundation of China (Grant No.2013M541792).

References

- [1] Wang Yuan-ming, ZHAO Dao-zhi, HUANG Jian 2009 The supply quality cost control model based on risk transmission *Science and Technology Management Research* **29**(3) 155-157
- [2] Tang Mei 2011 *Research on quality warranty contract in supply chain considering moral-hazard* Master's thesis, Nanjing University of Aeronautics and Astronautics
- [3] Zhou Yan-ju, Qiu Wan-hua, Wang Zong-run 2006 A Review on Supply Chain Risk Management *Systems Engineering* **24**(3) 1-7
- [4] Levy P 1998 *Total quality management in the supply chain. Hand book of TQM* Kluwer Academic, London 275-303
- [5] Chu-Hua Kuel, Madu C N 2003 *How to outperform Competitors with SCQM* The Lutin School of Business of Pace University
- [6] Svensson G 2000 A conceptual framework for the analysis of vulnerability in supply chains *International Journal of Physical Distribution & Logistics Management* **9**(30) 731-49
- [7] Tang C S 2006 Perspectives in supply chain risk management *International Journal of Production Economics* (10) 45 1-488
- [8] Batson R G, McGough K D 2007 A New Direction in Quality Engineering: Supply Chain Quality Modelling *European Journal of Operational Research* **45**(23) 5455-5464
- [9] Dahiya S, Khar R K, Chhikara A 2009 Opportunities, Challenges and Benefits of using HACCP as a Quality Risk Management Tool *The Pharmaceutical Industry Qual Assure* (12) 95-104
- [10] Zhou Chao-yun, Lu Zhi-qiang 2011 Analysis of Operation Decision in Supply Chain Based on Asymmetrical Quality Risk *Journal of Shanghai Jiao-tong University* **45**(12) 1782-1787
- [11] Ye Han-yan 2008 *Study on supply chain risk management and control* Ph.d. dissertation, Southwest Jiaotong university
- [12] Gu Li-gang, Gao Tao 2009 Study on quality risk disruption management in supply chain *Standard Science* (5) 4-7
- [13] Qian Ying 2007 *Study on coordination of supply chain quality management based on SCOR* Ph.d. dissertation, Hohai university
- [14] Yan Zhong-e 2013 The transfer model investigation of QFD-based supply chain quality risk *Science & Technology Progress and Policy* **30**(12) 22-25
- [15] Carol J R, Manoj K M 2008 Defining the concept of supply chain quality management and its relevance to academic and industrial practice *International Journal of Production Economics* (26) 315-337
- [16] Thomas S, Foster Jr 2008 Towards an understanding of supply chain quality management *Journal of Operational Management* (26) 461-467
- [17] Jiang Jia-dong, Zhao Han-ping, Feng Yun-cheng 2008 Study on the feature of quality risk in supply chain *Aeronautic standarzation & quality* (23) 29-33
- [18] Xu Fuyuan, He Jing.First 2002 Study on Supply and Demand Network with Multi-functional and Opening Characteristics for Enterprise *Forecasting* **21**(6) 19-22
- [19] Liu Cai-hong, Xu Fu-yuan 2008 Research on evolutionary game of SDN sub-network Systems engineering and electronics **30**(7) 1269-1272
- [20] Zhang Hai-feng 2004 Analysis on risk factors of strategic cooperation among supply chain enterprises *Journal of Wuhan Metallurgical Manager's Institute* **14**(4) 26-29
- [21] Xiao Li-min 2006 Empirical Research on Early Warning System of International Engineering Contract *Journal of Management Sciences* **19**(5) 75-82
- [22] Sr Shao-yong, chen Ji-ming, xu Dan, et.al. 2006 Agent-based Behaviour Modelling in Virtual Environment *Journal of System Simulation* **18**(1) 114-119
- [23] Zhang Hang-yi 2003 The Study About The CGF Action Modelling Technology Based On Agent *Computer Simulation* **20**(8) 79-81
- [24] Cai Yuan-li, Yu Zhen-hua, Zhang Xin-man 2007 Formal Modelling Methodology for Multi-Agent Systems *Journal of System Simulation* **19**(14) 3151—3157

Authors



Liu CaiHong, born in October, 1974, Hangzhou, Zhejiang Province

Current position, grades: Associate Professor within the School of Business in Jiaxing University and has a postdoctoral work in Zhejiang University.
Scientific interest: supply chain management, and quality engineer.

Experience: She is presided over the project supported by the natural science foundation of Zhejiang province in China (Grant No.LY13G010005) and the postdoctoral foundation of China (Grant No.2013M541792).



Xiong Wei, born in March, 1963, Hangzhou, Zhejiang Province

Current position, grades: Professor within the School of Management in the Zhejiang University.

Scientific interest: His researcher activity is in the area of quality engineer, and logistics.

Experience: He is presided over the project supported by the major program of the National Social Science Foundation (Grant No.12 &ZD206).

A study on the lock-in risk in IT outsourcing projects: the mechanism and the control system

Cong Guodong*

School of Business Administration, Zhejiang Gongshang University, Xuezheng Str.18, Hangzhou, P.R.China

Received 1 March 2014, www.tsi.lv

Abstract

This paper, proposes an insight into the mechanism of lock-in risk, which is the primary risk and the greatest concern in managing IT outsourcing projects. Based on Transaction Cost Theory (TCT), it develops an integrated three-layer model to instruct the mechanism of lock-in risk, namely, what will lead to lock-in risk (risk drivers), and what lock-in risk will result in (risk consequences). Four risk drivers such as asset specificity and three risk consequences such as cost escalation are identified and discussed in detail. Meanwhile, it instructs why four kinds of control could mitigate lock-in risk both from the sources and the consequences. Therefore, the effectiveness and reliability is enhanced throughout the risk management process of IT outsourcing projects, particularly for dynamic risk identification, controlling and monitoring.

Keywords: (project management, lock-in risk, IT outsourcing, risk management)

1 Introduction

With the rapid development of IT outsourcing projects, the accompanied high-risk is also highlighted. Since there exist broadly the information asymmetry throughout the outsourcing process, clients are inevitably faced with lock-in risk, even large companies have to suffer excessive dependence on vendors while outsourcing information system projects [1]. Lock-in is such a state that, without bearing the loss or sacrificing some or sometime even all of the assets to the vendor, the client can not get out of the relationship [2]. Lock-in state will result in such hazard as the technical capacity declining, or losing bargaining power, and cost escalation [3]. Therefore, lots of academic research has paid attention to lock-in risk with different perspective, such as Transaction Cost Theory (TCT) [4] and relationship governance, either focuses on service provider behaviour [5] or from the perspective of service buyer in large multinational companies [6]. The research either provides a framework or approach to assess lock-in risk.

However, little in-depth and specific analysis is proposed on the mechanism of lock-in risk, which leads to negative impact the effect of mitigation system. Especially in IT outsourcing projects, just some of the risk factors related to IT outsourcing projects are identified, let alone the factors are attained in certain countries within limited cultural contest [7]. This paper, addressing the issue, illustrates the mechanism of lock-in risk based on TCT. Namely, it constructs an integrated three layer model to demonstrate the transmission mechanism from four risk drivers (the set of risk factors)

to the risk, and then from the risk to the three risk consequences. Meanwhile, it proposes a clear instruction on why and how four kinds of control will carry out on the four risk drivers and three risk consequences, which improves the applicability and reliability of the analysis on the lock-in risk.

2 The mechanism of the lock-in risk

Hereinafter the mechanism of the lock-in risk will be demonstrated as a three-layer model, which is consisted of the transmission mechanism from four risk drivers (the set of risk factors) to the risk, and then from the risk to the three risk consequences.

2.1 THE MECHANISM BETWEEN RISK DRIVERS AND LOCK-IN RISK

The risk drivers are the reasons leading to lock-in risk. The definition, characteristics and function on lock-in risk are illustrated as below.

2.1.1 Small number of vendors

As IT outsourcing projects are closely integrated with organization management at all management levels, it is easier to form path dependence within IT outsourcing process. If clients choose only a small number of or even a single outsourcing service provider, they will trap themselves in lock-in risk.

The obvious situation is if their choice is limited, they will be under disadvantage in contract negotiation and

* Corresponding author - Tel: +86-15934214332; fax: +86-0571-28004233; E-mail: cgd@zjgsu.edu.cn

changing vendors [8]. Worse still, vendors are likely to adopt more opportunistic behaviour, e.g. to take the opportunity to bargain during the entire contract period or at updating occasion, as vendors are more aware of the true cost than other tenders.

Without exploiting multiple vendors to promote competition among vendors, there will be no convenience to evaluate a single vendor [2]. The other hazard is, without common vision and values, which will encourage both sides more commitment to partnership development, vendors will prohibit themselves to perform their duties with correct attitude throughout the outsourcing process.

Moreover, it is more difficult to execute control without adequate competition during the outsourcing process, especially because of the enhancement of vendors' advantageous position. Finally, clients have to be more aware of the importance to summarize experience and lessons, which will improve the capability to discover and correct potential opportunistic behaviour.

2.1.2 Asset specificity

Asset specificity, which concerns investments made specifically on physical or human resource because of a given contract. The asset is utilized in particular relationship, and need transfer cost if the utilization is changed. Additionally, the value of the relationship-specific investments would fall if the relationship dissolves [9]. The asset specificity IT outsourcing project vendor invest includes information system team members training and equipment, reallocation of human resources and improvement of information system development environment. Among the specific assets, vendor investment and human resource attribute more to the outsourcing risk than other asset [10].

To some extent, asset specificity enhances clients' concern on the loss or the sacrifice the asset invested in the relationship, accordingly results in lock-in risks. Especially, switching vendor costs have a greater impact on the lock-in problem. Having invested a great deal of time and effort in getting the initial vendor fully operational, the client itself may be reluctant to do so with a new vendor. Since some clients do not retain in-house competencies with the outsourced activity, they may even be unable to do so [8].

For experienced vendors, they are good at turning the specific investment asset into bargaining weapon with clients while updating contract. The reason is that, the specific investment asset becomes the entry barriers for other vendors, which means the other side in the outsourcing relationship has to invest at least the same amount to get the opportunity.

On the other hand, without any objective compatibility on the organization layer, both vendors and clients would not put more effort into long-term relationship. More importantly, they will not make use of the specific asset and reduce opportunistic behaviour. In other words, they will not enhance the bilateral condition

in the relationship and improve asset value, which will keep both sides in win-win situation. It is not the perspective that both clients and vendors are willing to see, even for the "worst" vendor that is fond of opportunistic behaviour, so both sides in the outsourcing relationship should strive for it.

2.1.3 Relatedness

Relatedness, also known as mutual dependence or connectivity, refers to mutual connection between tasks, business and functions [10]. The relatedness in IT outsourcing projects can be classified into two categories [11]: the relatedness between the business outsourced and stayed in-house, and that between the different kinds of business outsourced. In practice, two layers of relatedness should be distinguished, namely, the technical layer, which mainly focuses on the close connection among many kinds of technology adopted; the other is the management layer, which means the connection with internal management directly or indirectly, including the relatedness among the teamwork.

As proposed, risk factor 'Relatedness' has the lowest priority [7], it is effective "in Iranian organizations" though. The relatedness with the operation and the business enhances the reliance of clients on vendors. After an operation or a business is outsourced, if another operation and business closely related to the outsourced one, the original vendor owns a clear advantage over competitors, because it understands better than clients about the client's business, processes, style and even culture. Vendors may also take advantage of information asymmetry, or exaggerate the impact of the relatedness, so that clients have to rely on vendors' support not only in dealing with normal business, but also in the decision-making and management transformation.

Unless contract provides a benchmark for the boundary and responsibility of business units, the content for defining responsibility and compensation while service failure or loss occurs, therefore, the potential hazard of the relatedness will be out of control and the benefit of both sides will not be protected.

The most complicated problem of the relatedness is the controversial part of so-called 'extra work' or 'fuzzy work'. Unless there is a system to facilitate the climate of responsibility for vendors, so that they are willing to meet the requirement from client even if the requirement is beyond the contract strictly scope. Only in a high degree of trust, they would prefer resolving problems within reasonable scope to bargaining, since they are confident of better return in the long-term cooperation.

2.1.4 Client's expertise in IT operations

The term expertise is defined as special skill and knowledge from training, learning and practice [10], including abstract information, knowledge accumulation and skill in people's memory. In the IT outsourcing

project context, the definition is adopted that, expertise is the capability to combine external information and internal process to create particular ability. Risk factor “vendor’s lack of expertise with an IT operation” has been identified as the best risk factor [7], just as the author mentioned, that is “experts’ opinions in Iranian organizations”. From the perspective of clients, no other expertise is more important than that of their own, since that is the best equipment to protect their own benefit in the long process of IT outsourcing project.

Lack of expertise will lead to two kinds of hazard. The first one is that organization learning affects greatly on lock-in risk. The reason is, vendors occupy a comparative advantage about information, knowledge and skills in relative area, which forces clients to accumulate adequate technical expertise and knowledge. Otherwise, path dependence will push clients on the downstream of declining ability on information acquisition, data processing, independently identifying and solving problems, and even no intellectual property to govern vendors’ modifications in event of termination.

The other hazard is, even if vendors decline the service satisfaction for clients, clients will find it difficult to find out the truth for their unawareness of the expectation and need of users. Therefore, they lock themselves in the relationship without any opportunity to terminate the contract while vendors are under performance requirement, which demonstrates their vivid example of lock-in.

Without firm assurance between both sides, such as relational governance, it will be difficult to promote a beneficial atmosphere for strengthening organizational learning and cross-organizational knowledge transferring. Both client and vendors could not benefit from the trust between them: for clients, the efficiency of mastering the latest technology will not be improved, and the way to collaborate technology and business will not be found, which means the capacity of managing IT assets will not be enhanced; for vendors, better understanding on client industry and businesses will not be attained, and the vision of business will not be expanded. In other words, they are far from win-win. Undoubtedly, both sides are reluctant to lock themselves in such situation.

2.2 THE MECHANISM BETWEEN LOCK-IN RISK AND RISK CONSEQUENCES

Clients are aware of their dilemma of lock-in, and they are trying to alleviate the “painfulness”, such as form a well working relation with its ITO vendor for handling emergent issues, daily operation and events that have not been foreseen [6]. However, clients have to deal with three risk consequences as shown below:

2.2.1 Cost escalation

The first and the greatest, in most occasions, is the cost escalation. As long as clients are locked in, they will be

incapable of monitoring the scope and content of the service provided by vendors. Additionally, since they know less and less about the latest development of technologies related to the outsourced business, they will also lose the right to speak on what kind of technology could match and implement the business outsourced.

Therefore, vendors might attain great advantage of opportunistic behaviour. E.g. they may persuade clients into adopting extraordinary hardware, upgrading or developing more software. Even if the worst situation is not happening, namely, clients are confused and manipulated to pay for ‘useless’ service, the service cost will increase significantly.

2.2.2. The debasement of service quality

The second obvious risk consequence is the debasement of service quality. The reason is also mainly that, service providers’ behaviour is often on the edge of “opportunistic” [5]. Specifically, it is the vendors, rather than clients, that establish the evaluation criteria about service quality. Even clients are aware of problems or speak out of complaints, vendors are good at finding sufficient excuses to prove that they meet or even exceed the quality requirements of the contract, and then blame clients’ complaints to additional requirements, then seek to increase service charges for amending the contract.

2.2.3 The descending of client’s value

Another risk consequence need to be treated seriously is the descending of client’s enterprise value, including internal and external value. The internal value descending refers to the deterioration of competitiveness and the staff identification to the enterprise; the external value descending refers to the value impact outside of the enterprise, such as market value. If the enterprise is a listed company, the value will be directly reflected by stock price and market capitalization.

An open market is agile to the announcement of IT outsourcing projects. As proposed, the offshore outsourcing has a positive impact on the enterprise value, namely, capital market approves the value of outsourcing deals that contain high asset specificity, because the market believes that the research and development to match customer characteristics and other targeted contracts will help the company gain a strategic advantage or innovation assets [12].

The reaction of open market, as mentioned above, is definitely the “positive” part of the story. However, the clients in lock-in risk will be reluctant to find themselves blocked in the opposite situation, or even tragedy. As long as the investors sense their “painfulness”, what kind of choice they will make is not hard to imagine.

3 The control system on lock-in risk: why and how

The control system is to incentive individuals to work for fulfilling the organization's target [13]. It is suggested that, two formal control, behaviour control and output control, and two informal control, self-control and clan control [14] are available for a client. Therefore, clients are capable of making use of various control measures as a control portfolio accordingly, with the success of outsourcing project in mind.

3.1 THE CONTROL SYSTEM ON RISK DRIVERS

Hereinafter it will be demonstrated in detail why and how the control system affects on the risk drivers to alleviate the lock-in risk.

3.1.1. *Small number of vendors*

It is more difficult to implement control without adequate competition during the outsourcing process, especially because of the enhancement of vendors' advantageous position. Under such circumstance, informal control is preferred since it is easier to promote full participation of both sides.

Informal control will never be treated as a complementary system in the IT outsourcing projects. E.g., clan control is helpful to establish common vision and values, which will encourage both sides more commitment to partnership development. Particularly, trust between two sides will facilitate vendors to perform their duties with correct attitude throughout the outsourcing process.

Formal control is the assurance of informal control. E.g., contract could also become a useful control tool. In a detailed contract, e.g. SLA (Service Level Agreement), contingency plan, clause changing management, coordinating system and so on should be taken into consideration; flexible payment, terms of stage performance evaluation should also be integrated into contract, which ensures the incentive function of contract, both positively and negatively. Particularly, SLA can improve mutual trust. The advice is both sides should pay attention to change management, since changes pose negative impact on the output of mutual trust and commitment.

Additionally, clients should be more aware of the importance to summarize experience and lessons, which will improve the capability to discover and correct potential opportunistic behaviour.

3.1.2 *Asset specificity*

Control system can constrain vendor behaviour from both team and individual level. Particularly, output control, can be combined with the input-output ratio of assets to incentive vendors; clan system, with its cultural

advantage, will aid vendors maintain the devotion and even passion on client service and satisfaction.

Moreover, cost monitoring could be utilized as the way to control exclusive use of assets and resolve problems in the process, which is the best way to protect the asset. The first step of cost monitoring is to preset the milestone, and the second one is to measure mathematically the important and the relative business cost, and then defines a reasonable fluctuation range of cost [15].

3.1.3 *Relatedness*

The purpose of control system is to identify problems and reduce negative effects derived from the relatedness. With the complexity of the relatedness taken into account, it is reasonable to make use of multiple means of control. E.g. the implementation of output control will improve vendors' performance in the outsourcing business and related business, in other word, to ensure them to "do the right thing"; behaviour control, on the other hand, is to promote vendors to comply with the contract honestly and strictly, namely, ensure them to "do the thing right"; moreover, informal control, is to encourage vendors to provide satisfactory service, which can also be labelled as "do the thing with right attitude."

As a complementary control tool, contract provides a benchmark not only for the boundary and responsibility of business units, but also the content for defining responsibility and compensation while bifurcation or loss occurs, therefore, the potential hazard of the relatedness is under control and the benefit of both sides is protected.

Moreover, cost control will regulate a reasonable fluctuation range of cost, detect and properly handle unexpected cost caused by the relatedness to keep the total cost under control. Another option is the application of management mode different with in-sourced business, that is, to allow vendors to be more proactive and creative to solve problems rather than entangle in unnecessary details.

3.1.4 *Client's expertise in IT operations*

The most important role of control system to contribute for the client's expertise in IT operations is the potential to encourage organizational learning and cross-organizational knowledge transferring. Both client and vendors benefit from the beneficial atmosphere between them: for clients, the willingness and the efficiency of following the latest technology is improved, and the way to collaborate technology and business is found, which means the capacity of managing IT assets is enhanced; for vendors, better understanding on client industry and businesses is attained, and the vision of business is expanded.

Informal control will contribute for the degree of expertise in IT operations, that is, establish the consistency of the interests for both sides, which reduces

the barrier of knowledge transferring. Under the framework of informal control, both sides are willing to employ flexible approaches such as cross-organization teams, virtual organizations to enhance the intellectual capital investment, and organizational learning, which eventually promote the orderly transformation from individual knowledge into organizational capacity.

Formal control could also utilize the standardization, namely, both sides adopt the same standard technical and professional language, which strengthens the understanding and communication, and continuously improves learning efficiency and knowledge quality. In this way, the degree of expertise in IT operations will be upgraded constantly for both client and vendors.

3.2 THE CONTROL SYSTEM ON RISK CONSEQUENCES

Hereinafter it will be demonstrated in detail why and how the control system affects on the risk consequences to prevent the lock-in risk.

If the risk is characterized by “potential” and “possible” hazards, risk consequence will be the “fact” of damage that has really happened and demonstrated. Hence, after the risk consequences occurred, if the source of the consequences could be traced and control measures could be carried out, the problems and the risks will be solved fundamentally. However, the other way should never be ignored, that is the direct control executed on the consequences themselves. The reason is, doing so is “the pill right to cure symptom” to the case, which is easy for both sides to understand and support.

3.2.1 The principle to control risk consequences

As for lock-in risk , the principle to control is: once one of the three risk consequences appeared, the first thing to do is to track the four risk drivers based on risk mechanism discussed above, then analyse the role of each risk drivers, and take the appropriate measures; secondly, apply control system on risk the consequences. Only in a parallel way, the objectives of lock-in risk management could be achieved effectively.

There exist lots of risks that lead to three risk consequences, and then will it be “unfair” to just blame lock-in risk? The answer is: as the biggest risk, it is always “safe” to probe into the issue from lock-in risk, and fortunately great harvest will seldom disappoint those staring at it.

In addition, the risk consequences caused by the lock-in risk also possesses distinctive features: the speed of problem deterioration is “stunning”, while clients are able to do nothing, or are left no other choice but allow vendors creating various excuses. These features are all due to the special status of lock-in risk itself, without any exception. E. g., cost escalation becomes a black hole, users keep complaining about service quality to nowhere,

employee loyalty has fallen sharply, and many negative comments spread on the internet, and so on.

Throughout the management process of lock-in risk, Iceberg Theory must be kept in mind: Behind every big problem, there will be at least thirty small hidden problems. Therefore, the control of lock-in risk should be executed from both the sources and the consequences, so that the best control results could be satisfied.

3.2.2 The measures to control risk consequences

Once any of the indications mentioned above emerges, the signal of risk consequences is alerting that control measures must be taken. Above all, clients will learn to take advantage of utilizing the control portfolio, namely, adopt one or some of the four controls according to the situation.

The implementation of output control will improve vendors’ performance in monitoring cost; behaviour control, on the other hand, is to promote them to keep an eye on the cost honestly and strictly; informal control, is to encourage them to provide extra effort to work out more effective ways to help clients prevent the cost out of control.

Take the control of cost escalation as an example. Clients could make out the benchmark of the service output and behaviour, namely, establish the regulation for a reasonable fluctuation range of cost, and then monitor the abnormal change in order not to fall into the trap of unexpected cost.

It is more challenging and promising to utilize informal control, yet there are several options by the best exercises and samples from the leading companies. E.g., one option is to apply a different management style in the outsourced business in order to allow vendors to be more proactive and creative to solve problems rather than entangle in unnecessary details. Another option is for CIO, namely, play a new role as “CIO leader” that emphasizes on strategic business issues of both sides and coordinate IT with business. In this way, IT outsourcing projects will be more powerful support for the emprise strategy, and will create more value for both sides than before.

In order to disclose the mechanism and the control system of the lock-in risk in IT outsourcing, this paper proposes an integrated three-layer model as shown in Figure 1.

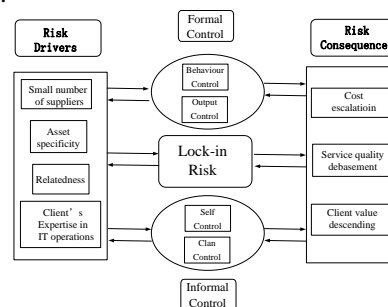


FIGURE 1 The three-layer model of lock-in risk

4 Summarization

This paper, intends to provide an insight into the mechanism of lock-in risk, which is the primary risk and the greatest concern in managing IT outsourcing project risk. Based on Transaction Cost Theory, it instructs the mechanism of lock-in risk with a three-layer model. Moreover, it proposes a clear instruction on why and how four kinds of control will carry out on the four risk drivers and three risk consequences, which improves the applicability and reliability of the analysis on the lock-in risk.

With the effort, the lock-in risk could be diagnosed and controlled systematically. In other words, the risk could be prevented beforehand, and also be tracked as long as risk consequences emerge. The paper contributes for both academic and practice as a helpful tool throughout the whole risk management process, especially for risk identification, evaluation and monitoring.

Future research will focus on the empirical method in order to further illustrate the contribution of each factor within the three-layer model of the lock-in risk, and further analyses the mutual relationship between certain factors. In addition, it will also study the effectiveness of mitigation system, namely, how the mitigation system work on the risk drivers rather than lock-in risk itself. Therefore, a complete management framework will be established both for academic and for practice.

Acknowledgments

This research is supported by Humanity and Sociology Foundation of Ministry of Education of China (Grant No. 10YJC630034), Zhejiang Provincial Natural Science Foundation of China (No. Y6110539).

References

- [1] Gonzalez R, Gasco J, Lopis J 2005 *Industrial Management and Data Systems* **105**(1) 45-62
- [2] Aubert B A, Rivard S, Patry M 2004 *Information and Management* **41** 921-32.
- [3] Bahli B, Rivard S 2013 *Decision Support Systems* **56** 37-47
- [4] Alagheband F K, Rivard S, Wu S, Goyette S 2011 *The Journal of Strategic Information Systems* **20**(2) 125-38
- [5] Mathew S K 2011 *Strategic Outsourcing: An International Journal* **4**(2) 179-200.
- [6] Hodosi G, Rusu L, Choo S 2012 *International Journal of Social and Organizational Dynamics in IT* **2**(3) 29-47
- [7] Keramati A, Samadi H, Nazari-Shirkouhi S 2013 *International Journal of Business Information Systems* **12**(2) 210-42
- [8] Aubert B A, Patry M, Rivard S. *Wirtschafts informatik* 2003 **45**(2) 181-90
- [9] Williamson O E 1985 *The Economic Institutions of Capitalism* The Free Press: New York chapter1
- [10] Bahli B, Rivard S 2005 *Omega* **33** 175-87
- [11] Earl M J 1996 *Sloan management review* **3** 26-32
- [12] Florin J, Bradford M, Pagach D 2005 *The Journal of High Technology Management Research* **16**(2) 241-53
- [13] Kirsch L J, Sambamurthy V, Dong-Gil K, Russell L P 2002 *Management Science* **48**(4) 484-98
- [14] Ouchi W G 1980 *Administrative Science Quarterly* **25**(1) 129-41
- [15] Osei-Bryson K M, Ngwenyama O K 2006 *European Journal of Operational Research* **174** 245-64

Authors



Guodong Cong, born in August, 22, 1972, Weihai, P.R.China

Current position, grades: lecturer and researcher

University studies: Zhejiang Gongshang University

Scientific interest: project management, IT outsourcing risk management

Publications: 9

Experience: Dr. Guodong Cong got his Ph.D. degree in Management from Huazhong University of Science and Technology, P.R. China. Before his Ph.D., he worked in a foreign trading company as a project manager, served for over ten IT projects as team leader and project manager. He also provided consulting and training service for over twenty famous companies on IT outsourcing projects. His email is cgd@zjgsu.edu.cn.

A game theory analysis - dumping by multinational company and antidumping in China

Bingjie Li*

School of Economics, Xi'an University of Finance and Economics, Xi'an, 710065, China

Received 1 March 2014, www.tsi.lv

Abstract

Based on a dynamic game model, this paper analyses dumping by multinational firm and Chinese government's antidumping behaviour. It is shown that no matter how many products are dumped in domestic market; Chinese authority should impose punitive damages against the foreign firm as long as the scale of antidumping duty is not too high to stop multinational company's investment. This strategy will improve social welfare.

Keywords: Multinational company, Dumping, Antidumping, Game

1 Introduction

Foreign capital inflows have been considered as an engine for Chinese economic growth since China's reform and open [1]. The capacity and sale of multinational companies turn to be very large scale. At the same time, the indisputable fact is that multinational companies dump their products in not only Chinese traditional industries but also the emerging industry [2]. There was dumping by foreign firm in many industries like steel, chemical, oil, machine tool, plywood, computer, food and beverages. This is inevitably harmful to the development of relevant industries in China.

It is worth mentioning that the form of dumping is not confined to regular dumping but also pre sales, a foreign investment preferential policy provided by Chinese government. Pre-sales means the foreign companies introduce products, which are produced in other market into the host country before their capacity in the host country goes into operation. In this way, foreign firms can expand their market shares thereafter. The penetration pricing method is broadly adopted by multinational companies in the early stage of the market development (namely a way of pricing marketing in which foreign firms intentionally sell a product at a lower price to stimulate the market demand and improve the market share) [3]. However, this way of selling and pricing is a typical form of dumping and becomes the focus of the Chinese government import anti-dumping.

By the end of March 2011, China has launched 189 cases of anti-dumping investigation on imported products, 177 cases of which are filed after China's access to the WTO [4]. In ten years, China has become one of the countries that implement anti-dumping investigations most frequently. In this period, 26

countries and regions were involved in anti-dumping investigations by China's authority. United States and Japan attracted the majority of antidumping suits. Their combined share accounted for up to 48%. South Korea was in the second place, about 46% [5]. The third is the European Union, about 25%. It is not hard to see the countries or regions that were ruled guilty of dumping mainly are developed countries, which have industrial strength and high responding ability.

The foreign firms engaged in dumping mostly concentrate in nine industries such as chemicals, steel, metalworking, automobile, mechanical and electrical products, papermaking, textile and food [6]. This is relevant to the development tendency of China's internal and external trade policy after China became a member of the WTO. Although chemical industry and steel metallurgy industry attracted significant foreign investment, there were the most number of anti-dumping appeals [7].

2 The Setup

2.1 ASSUMPTION

We use dynamic game theory to analyse the foreign enterprise's use of pre-sales for dumping and anti-dumping measures taken by the Chinese government, and make the following assumptions [8]:

- (1) The initial market structure of the industry is a symmetry oligopoly. There are static or dynamic economies of scale, i.e. each firm in the game has a decreasing marginal cost curve.
- (2) Government provides certain preferential policies to foreign enterprises, which makes difference in the cost functions of foreign and Chinese firm.

* Corresponding author - Tel: +86-132-597-68686; Fax:+86-029-8540-3156; E-mail: muzi6shui@126.com

Assumption (1) and (2) imply that average cost of foreign companies is lower than that of Chinese company because the foreign firm is relatively larger than Chinese firm and can enjoy the preferential policies only available to alien firm.

(3) Firms are producing a homogenous good. Hence, the products of foreign firm and Chinese firm are perfect substitutable.

(4) Foreign firms and Chinese companies engage in Cournot competition and will not quit the game in the $[1, t]$ period since they are pursuing long-term strategic interests.

Moreover, let T_p be a dummy variable and p be the time. $T_p = 1$ indicates there is dumping by foreign firm in pre-sales period (p) and in ($p+1$) period Chinese government imposes anti-dumping duty which equals the margin of dumping. $T_p = 0$ means there is dumping by foreign firm in pre-sales (p) period without anti-dumping charge from the Chinese government in period ($p+1$). The multinational's and the domestic firms' profits (W) are then:

$$W_{p,h} = [P_{p,h}(Q_p) - C_{p,h}]Q_{p,h}, \tag{1}$$

$$W_{p,f} = [P_{p,h}(Q_p) - C_{p,f} - T_{p-1}S_{p-1}]Q_{p,f}, \tag{2}$$

where C is cost and Q is quantity. S_{p-1} denotes the margin of dumping and $P_{p,h}(Q_p)$ is price in Chinese market.

2.2 ANTIDUMPING STRATEGY BY CHINESE AUTHORITY

Profit maximizing with respect to quantity gives out the first order condition as following:

$$\partial W_{p,h} / \partial Q_{p,h} = P_{p,h} + P'_{p,h}Q_{p,h} - C_{p,h} = 0, \tag{3}$$

$$\partial W_{p,f} / \partial Q_{p,f} = P_{p,h} - C_{p,f} - T_{p-1}S_{p-1} + P'_{p,h}Q_{p,f} = 0. \tag{4}$$

As it is perfect substitutive between foreign product and domestic product, the following condition needs to be satisfied:

$$\begin{aligned} \partial^2 W_{p,h} / \partial Q_{p,h} \partial Q_{p,f} < 0, \\ \partial^2 W_{p,f} / \partial Q_{p,h} \partial Q_{p,f} < 0, \quad \partial^2 W_{p,h} / \partial Q^2_{p,h} < 0, \\ \partial^2 W_{p,f} / \partial Q^2_{p,f} < 0. \end{aligned}$$

Then

$$\begin{cases} \partial^2 W_{p,h} / \partial Q^2_{p,h} = 2P'_{p,h} + P''_{p,h}Q_{p,h} < 0 \\ \partial^2 W_{p,f} / \partial Q^2_{p,f} = 2P'_{p,f} + P''_{p,h}Q_{p,h} < 0 \\ \partial^2 W_{p,h} / \partial Q_{p,h} \partial Q_{p,f} = \partial^2 W_{p,f} / \partial Q_{p,h} \partial Q_{p,f} = 2P'_{p,h} + P''_{p,h}Q_{p,f} < 0 \\ \partial^2 W_{p,h} / \partial Q^2_{p,h} < \partial^2 Q_{p,h} / \partial Q_{p,h} \partial Q_{p,f} \\ \partial^2 W_{p,f} / \partial Q^2_{p,f} < \partial^2 Q_{p,f} / \partial Q_{p,f} \partial Q_{p,h} \end{cases} \tag{5}$$

$$H_p = \det \begin{bmatrix} \partial^2 W_{p,h} / \partial Q^2_{p,h} & \partial^2 W_{p,h} / \partial Q_{p,h} \partial Q_{p,f} \\ \partial^2 W_{p,f} / \partial Q_{p,f} \partial Q_{p,h} & \partial^2 W_{p,f} / \partial Q^2_{p,f} \end{bmatrix} < 0 \tag{6}$$

The unique Cournot-Nash equilibrium from equations (1)-(4) can be labelled as $P_{p,h}^{CN}(T_{p-1}S_{p-1})$ and $Q_{p,h}^{CN}(T_{p-1}S_{p-1}), Q_{p,f}^{CN}(T_{p-1}S_{p-1})$. If foreign firm is dumping in the pre-sale period ($p-1$), Chinese government imposes anti-dumping duty, namely $T_{p-1} = 1$, derivative of equations (3)-(4) with respect to Cramer's rule will give the following formula:

$$\begin{cases} \partial Q_{p,h}^{CN} / \partial S_{p-1} = -P'_{p,h} (P'_{p,h} + P''_{p,h}Q_{p,h}) / H_p > 0 \\ \partial Q_{p,f}^{CN} / \partial S_{p-1} = P'_{p,h} (3P'_{p,h} + P''_{p,h}Q_{p,h}) / H_p < 0 \\ \partial P_{p,h}^{CN} / \partial S_{p-1} = P'_{p,h} \bullet \partial Q_{p,h}^{CN} / \partial S_{p-1} > 0 \end{cases} \tag{7}$$

As a result, we obtain Proposition 1.

Proposition 1: When there is dumping by foreign firm in pre-sale period ($p-1$), and Chinese government issues an antidumping duty order to offset the injury, the international market prices will rise in (p) period, and foreign country's exports decline, but Chinese domestic enterprises will expand production.

Partial derivative of equilibrium profits with respect to S_{p-1} obtain the following inequality

$$\begin{cases} \partial W_{p,h}^{CN} / \partial S_{p-1} = (P'_{p,h})^2 \bullet Q_{p,h}^{CN} \bullet (2P'_{p,h} + P''_{p,h}Q_{p,h}) / H_p > 0 \\ \partial W_{p,f}^{CN} / \partial S_{p-1} = -Q_{p,f}^{CN} [1 + (P'_{p,h})^2 \bullet (2P'_{p,h} + P''_{p,h}Q_{p,h})] H_p < 0 \end{cases} \tag{8}$$

Long-term profit maximization is defined as:

$$\begin{cases} \max W_h = W_{p-1,h} + \tau \xi \partial W_{p,h}^{CN} + \tau(1-\xi)W_{p,h}^{CN} (0) \\ \max W_f = W_{p-1,f} + \tau \xi \partial W_{p,f}^{CN} + \tau(1-\xi)W_{p,f}^{CN} (0) + \\ W_{p-1,0}(L_{p-1}) + \tau W_{p,0}(L_p) \end{cases} \tag{9}$$

where $W_{p,0}$ and L_p denote profit and sales of foreign firm in home country, respectively while ξ is probability that Chinese government issues an anti-dumping duty order in pre-sale period.

The equilibrium output can be determined by solving the first-order condition for equation (8):

$$\begin{cases} \partial W_h / \partial Q_{p-1,h} = \partial W_{p-1,h} / \partial Q_{p-1,h} + \tau \xi \partial W_{p,h}^{CN} / \partial S_{p-1} \bullet \partial S_{p-1} / \partial Q_{p-1,h} = 0 \\ \partial W_f / \partial Q_{p-1,f} = \partial W_{p-1,f} / \partial Q_{p-1,f} + \tau \xi \partial W_{p,f}^{CN} / \partial S_{p-1} \bullet \partial S_{p-1} / \partial Q_{p-1,f} = 0 \\ \partial W_f / \partial Q_{p-1} = \partial W_{p-1,0} / \partial L_{p-1} + \tau \xi \partial W_{p,h}^{CN} / \partial S_{p-1} \bullet \partial S_{p-1} / \partial L_{p-1} = 0 \end{cases} \quad (10)$$

Since $\partial S_{p-1} / \partial Q_{p-1,h} = -P'_{p-1,h} > 0$, $\partial W_{p,h}^{CN} / \partial S_{p-1} > 0$, equation (10) implies $\partial W_{p-1,h} / \partial Q_{p-1,h} < 0$. It means Chinese firm's profit will fall when foreign company dump in China and will rise when Chinese government imposes anti-dumping duty.

If there is dumping by foreign firm in pre-sale period, i.e. $S_{p-1} > 0$, with $\partial S_{p-1} / \partial L_{p-1} = P'_{p-1,h} < 0$ and $\partial W_{p,f}^{CN} / \partial S_{p-1} < 0$, the following inequality satisfies $\begin{cases} \partial W_{p-1,f} / \partial Q_{p-1,f} > 0 \\ \partial W_{p-1,0} / \partial L_{p-1} < 0 \end{cases}$, which imply that dumping by foreign company will make its export earnings rise, but profit of foreign firm in its home country will decline. As a result, we have Proposition 2 and Proposition 3.

Proposition 2: The profit of parent company in the home country is decreasing in the anti-dumping duty imposed on their affiliate in China in the pre-sales period. While the profit of Chinese domestic enterprises is increasing in the anti-dumping tax, but the enthusiasm of foreign investment may be affected.

How many anti-dumping duties should the Chinese government to impose? It should have warning and sanction effect on foreign companies but should not affect their investment enthusiasm. We will turn to this issue later.

Proposition 3: The aim of Chinese government to impose anti-dumping duty is to encourage domestic enterprises to increase output and warrant fair competition. Anti-dumping is a threaten action ex ante and has sanction effect ex post.

Let U denote utility function. Because anti-dumping tax is levied on Chinese firms that import the product involving dumping, it is not relevant to national welfare. Specifically, China's social welfare is defined as:

$$F_{p,h} = U(Q_p) - P_{p,h}Q_p + (P_{p,h} - C_{p,h})Q_{p,h} \quad (11)$$

Derivative with respect to S_{p-1} is given by:

$$dF_{p,h} / dS_{p-1} = (P'_{p,h})^2 [Q_p [P'_{p,h} - \Psi_{p,h} (2P'_{p,h} + P''_{p,h} Q_{p,h})] / H_p] \quad (12)$$

where $\Psi_{p,h} = Q_{p,h} / Q_p$ is the market share of Chinese firm in t period.

Proposition 4: If $\Psi_{p,h} > (= or <) P'_{p,h} / (2P'_{p,h} + P''_{p,h} Q_{p,h})$, anti-dumping policy by Chinese government will increase (not change or decrease) the domestic social welfare. The greater the dumping margin, the better (or worse) China's social welfare will become.

If $\Psi_{p,h} \rightarrow 1$, i.e. import volume is trivial, as $dF_{p,h} / dS_{p-1}$ is proportional to $dQ_{p,h} / dS_{p-1}$, China's national welfare still can be better off when the authority take anti-dumping measures.

It implies that it is necessary for China's government to fight unfair competition from imported product, no matter how small the amount of dumping.

2.3 THE SCALE OF ANTIDUMPING DUTY

The effect of industrial preferential policies and anti-dumping tax on cost of foreign firms are denoted by μ and λ , respectively, which are distributed according to a continuous 0 mean distribution. $\mu > 0$ represents that cost of a foreign company is superior to that of a Chinese firm; $\lambda > 0$ implies that foreign firms are levied heavier tax burden than Chinese firms. According to the former definition about foreign investment policies and taxing policies, we denote $\alpha\lambda$ the effect of anti-dumping on unit cost of a Chinese firm. Then we specify the profit function of a foreign firm and a Chinese firm as follows:

$$\begin{aligned} w_h &= \sum_{i=1}^{\alpha} [\pi_i(x)] = \sum_{i=1}^{\alpha} [x_i p(x) - c_i(x_i) - \mu x_i + \alpha \lambda x_i] \quad (13) \\ &= x_h p(x) - c_h(x_h) - \mu x_h + \alpha \lambda x_h \end{aligned}$$

$$\begin{aligned} w_f &= \sum_{j=1}^{\beta} [\pi_j(x)] = \sum_{j=1}^{\beta} [x_j p(x) - c_j(x_j) - \mu x_j - \lambda x_j] \quad (14) \\ &= x_f p(x) - c_f(x_f) + \mu x_f - \lambda x_f \end{aligned}$$

where $\pi_i(x)$ is the profit function of China's domestic firm and $\pi_j(x)$ is the profit function of foreign firm. Profits depend on output x_1, x_2, \dots, x_n , inverse demand function $p(x)$ and total production cost $c_h(x_h), c_f(x_f)$.

Specifically, we define $p_i(t) = \frac{1}{\alpha e^{\beta_i t}}$, where t is time node and β_i is the price of the i-th product, as a result $p(x) = \int_0^{\infty} p_i(t) = \int_0^{\infty} \frac{1}{\alpha e^{\beta_i t}}$. In addition, production costs satisfy $C'_h, C'_f > 0, C''_h, C''_f < 0$.

Substituting $p(x) = \int_0^{\infty} p_i(t) = \int_0^{\infty} \frac{1}{\alpha e^{\beta_i t}}$ into equations (13)-(14) yields the competitors' profit maximization function:

$$\begin{aligned} \max W_h &= \max \sum_{i=1}^{\alpha} [\pi_i(x)] \quad (15) \\ &= \max [x_h \int_0^{\infty} \frac{1}{\alpha e^{\beta_i t}} - c_h(x_h) - \mu x_h + \alpha \lambda x_h] \end{aligned}$$

$$\begin{aligned} \max W_f &= \max \sum_{j=1}^{\beta} [\pi_j(x)] \\ &= \max [x_f \int_0^{\infty} \frac{1}{\alpha e^{\beta t}} - c_f(x_f) - \mu x_f - \alpha x_f] \end{aligned} \tag{16}$$

If dumping by foreign firms is imposed anti-dumping duties, there is investigation cost and other relative expenditure. We denote by γ this kind of cost satisfying $0 < \gamma < 1$. Therefore the effect of anti-dumping on unit cost of a Chinese firm changes from $\alpha\lambda x_h$ to $\alpha\gamma\lambda x_h$. Partial derivative yields the competitors' first-order conditions:

$$\begin{cases} \frac{\partial w_h}{\partial w_f} = p(x) + x_h p'(x) - c'_h(x_h) - \mu + \alpha\gamma\lambda = 0 \\ \frac{\partial w_f}{\partial w_h} = p(x) + x_f p'(x) - c'_f(x_f) + \mu - \lambda = 0 \end{cases} \tag{17}$$

Solving the first order condition gives the slope of respond function:

$$\begin{cases} k_h = - \frac{\partial^2 w_h / \partial w_f^2}{\partial^2 w_h / \partial w_h \partial w_f} \\ k_f = - \frac{\partial^2 w_f / \partial w_h^2}{\partial^2 w_f / \partial w_h \partial w_f} \end{cases} \tag{18}$$

The inverse demand function indicates that $\frac{\partial^2 w_h}{\partial w_h \partial w_f} < 0$; $\frac{\partial^2 w_f}{\partial w_h \partial w_f} < 0$ and $\frac{\partial^2 w_h}{\partial x_h^2} > \frac{\partial^2 w_h}{\partial x_h \partial x_f}$; $\frac{\partial^2 w_f}{\partial x_f^2} > \frac{\partial^2 w_f}{\partial x_f \partial x_h}$.

Therefore, $k_h < 0$, $k_f < 0$ and $|k_h| \geq |k_f|$. The equilibrium is $|k_h| = |k_f|$.

According to equations (17)-(18), we solve the equilibrium $\lambda = p\alpha/k$. This is the threshold condition for government to impose anti-dumping duties.

Consequently, the structure of the game is outlined in Figure 1, where foreign firm is player 1 and Chinese government is player 2. In the first stage, the foreign firm decides whether to dump in Chinese market. In the second stage, the Chinese government can react to the foreign firm's dumping decision in the first stage. That is whether to impose anti-dumping duty on foreign firm. In the third stage of the game, foreign firms decide either to withdrawal investment from China or keep investing in China but stop dumping.

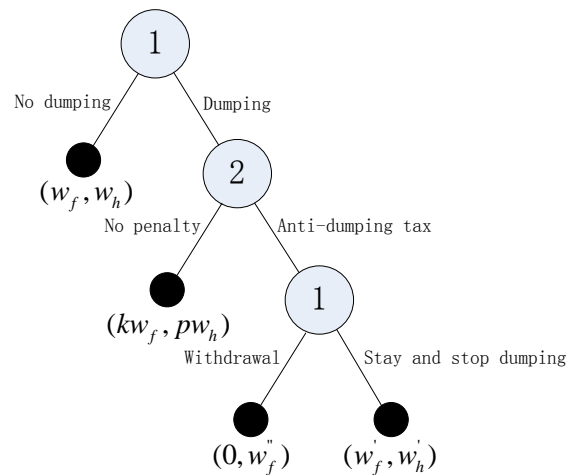


FIGURE1 Game Structure

It is clear from the game tree that the sum of profits of foreign and Chinese firm is maximized, when foreign firm does not dump and when $|k_h| = |k_f|$. In this case, the Chinese firm yields the most profit. Hence, China's government, by making a few laws and regulations to improve market competition and alerts the foreign enterprise, can optimize the total profit and ensure the interests of Chinese enterprises as well. If the foreign company chooses to dump product into Chinese market, its profit will increase to kw_f , but the Chinese firm's profit decrease to pw_h . Then the overall loss of profit is $(1-p)w_k - (k-1)w_f > 0$. In this case, if Chinese government does not intervene, eventually dumping by foreign firms will force Chinese companies out of the market. In case of intervention by imposing anti-dumping duty and providing certain subsidy to Chinese firms, the sum of profit of both kind of firm is $(kw_f - \lambda x_f) - w_f + w_h - (pw_h + \alpha\lambda x_h)$.

If $\lambda < p\alpha/k$, the above overall profit is minus but the foreign firm is still profitable through dumping. So, $\lambda < p\alpha/k$ is the threshold condition for government to impose anti-dumping duty on foreign firms. If the anti-dumping duty is too high, making the foreign capital withdraw from the Chinese market and Chinese companies' profit increase, the market cost and other relevant cost will increase, and because of the destruction of the market structure caused by the previous price war, overall market profits will become $w_h'' < w_h + w_f$. It will lead to Nash equilibrium again in the event that foreign enterprise stop dumping because of anti-dumping tax and stay in China. At this moment, although market is distorted to some extent and profits of foreign and Chinese firm both decline, the overall profit still optimize as $w_f' + w_h'$.

3 Conclusions

Multinational companies play a main role in FDI inflow to China and are influenced by preferential investment policy provided by Chinese government. Theoretically, speaking, the conflict between anti-dumping and investment absorption stem from the clash of interest of Chinese government and multinational companies. The aim for multinational companies to invest internationally is to strengthen their global market competitiveness, so as to realize their own profit maximization, while the Chinese government concerns about promoting China's economic development and enhancing the global competitiveness of Chinese enterprises, so as to improve social welfare.

Based on a dynamic game model, we draw some conclusions:

(1) When there is dumping by foreign firm and Chinese government issues an anti-dumping duty order to offset the dumping, the international market price will rise and foreign firm's exports from its home country decline, but Chinese domestic enterprise will expand its production.

(2) The profit of the parent company in the home country is decreasing in the anti-dumping duty imposed on its affiliate in China in the pre-sales period, while the profit of Chinese domestic enterprise is increasing in the anti-dumping tax.

(3) The aim of Chinese government to impose anti-dumping duty is to encourage domestic enterprises to increase output and warrant fair competition. Anti-dumping is a threaten action ex ante and has sanction effect ex post.

(4) If $\Psi_{p,h} > (= \text{or} <) P'_{p,h} / (2P'_{p,h} + P^*_{p,h} Q_{p,h})$, anti-dumping policy by Chinese government will increase (not change or decrease) the domestic social welfare. The greater the dumping margin, the better (worse) China's

social welfare will become.

If $\Psi_{p,h} \rightarrow 1$, i.e. import volume is trivial, as $dF_{p,h} / dS_{p-1}$ is proportional to $dQ_{p,h} / dS_{p-1}$, China's national welfare still can be better off when the authority take anti-dumping measures. It implies that it is necessary for China's government to fight unfair competition from imported product, no matter how small the amount of dumping.

(5) A proper scale of antidumping duty will not change the mode of entry by the multinational company. There will still be a great amount of FDI inflow.

Acknowledgments

In the process of writing paper, thanks for the help and guidance of my colleagues and supervisor.

References

- [1] Xi P 2006 Review and forecast about China foreign capital utilization *Foreign Investment in China* **2006**(7) 20-3 (in Chinese)
- [2] Gong J Y, Teng Y H 2005 Analysis of the effect of investment in the Implementation of Chinese anti-dumping *East China Economic Management* **19**(8) 72-5 (in Chinese)
- [3] Guo S T 2006 The export enterprises strategy analysis under anti-dumping conditions *Finance & Trade Economics* **2006**(3) 83-6 (in Chinese)
- [4] Yang Y 2009 *Anti dumping price effect and decision system* Beijing: Communication University of China Press (in Chinese)
- [5] Hu M X, Zhou Y Y 2005 Study on the foreign direct investment of antidumping jumping *Journal of Ningxia University* **2005**(3) 87-90 (in Chinese)
- [6] Tang Y 2004 Analysis of the four economic effect induced by antidumping protection *Finance & Trade Economics* **2004**(11) 65-9 (in Chinese)
- [7] Park S 2009 The Trade Depressing and Trade Diversion Effects of Antidumping actions: The Case of China *China Economic Review* **20**(3) 542-8
- [8] Zhang W Y 2002 *Game theory and information economics* Shanghai: Shanghai People's Publishing Press (in Chinese)

Author



Bingjie Li, born in March, 1977, Xian City, Shaanxi Province, P.R. China

Current position, grades: the lecturer of Faculty of Economics, Xian University of Finance and Economics.

University studies: PhD of Faculty of Economics and Finance, Xian Jiaotong University.

Scientific interest: International Trade, Financial and fiscal Management.

Publications: Presided over 3 scientific research projects the completion of provincial; more than 10 papers published in various journals.

Experience: Graduated from Xian Jiaotong University in 2008, has completed 3 scientific research projects; more than 10 papers published in various journals.

A new method based on induced aggregation operator and distance measures for fuzzy decision-making

Zhihong Ma¹, Jianping Chen^{2*}, Shouzhen Zeng³

¹Basic Science Department, Tianjin Agricultural University, Tianjin, 300384, China;

²School of Management, Ningbo Institute of Technology, Zhejiang University, Ningbo, 310018, China

³College of Computer and Information, Zhejiang Wanli University, Ningbo, 310015, China

Received 1 March 2014, www.tsi.lv

Abstract

In this paper we introduce a new fuzzy decision making model that unifies induced aggregation operators and distance measures in the same formulation. We develop the fuzzy induced generalized ordered weighted averaging distance (FIGOWAD) operator. The main advantage of this operator is that it provides a parameterized family of aggregation operators between the minimum and the maximum and a wide range of special cases. Another advantage is that it is able to deal with the fuzzy environment where the information is very imprecise and can be assessed with interval numbers. Moreover, it uses induced aggregation operators that provide a more general representation of the attitudinal character of the decision-maker. We study some families of the FIGOWAD operators. We end the paper with an application of the new approach in a business decision-making problem about the selection of strategies.

Keywords: Fuzzy numbers, induced aggregation operator, distance measure, decision making

1 Introduction

Different types of aggregation operators are found in the literature for aggregating the information [1-3]. A very common aggregation method is the ordered weighted averaging (OWA) operator introduced by Yager [4], whose prominent characteristic is the reordering step. The OWA operator provides a parameterized family of aggregation operators that includes as special cases the maximum, the minimum and the average criteria. Since its appearance, the OWA operator has been used in a wide range of applications such as [5-24].

An interesting extension of the OWA operator is the induced OWA (IOWA) operator [25, 26]. The IOWA operator differs in that the reordering step is not developed with the values of the arguments but can be induced by another mechanism such that the ordered position of the arguments depends upon the values of their associated order-inducing variables. The IOWA operator has been studied by different authors in recent years [27-38].

A further interesting extension is the one that uses the OWA and the IOWA operator in distance measures. Recently, motivated by the idea of the OWA operator, Xu and Chen [39] defined the ordered weighted distance (OWD) measure whose prominent characteristic is that they can alleviate (or intensify) the influence of unduly large (or small) deviations on the aggregation results by assigning them low (or high) weights. Yager generalized

Xu and Chen's distance measures and provided a variety of ordered weighted averaging norms, based on which he proposed several similarity measures between fuzzy sets. Merigó and Gil-Lafuente [40] introduced a new index for decision-making using the OWA operator to calculate Hamming distance called the ordered weighted averaging distance (OWAD) operator, and gave its application in the selection of financial products and sport management. Zeng and Su [41] and Zeng [42] extended Xu and Chen's result to intuitionistic fuzzy environment and presented the intuitionistic fuzzy ordered weighted distance (IFOWD) operator.

On the basis of the idea of the IOWA operator, Merigó and Casanovas [43] presented an induced ordered weighted averaging distance (IOWAD) operator that extends the OWA operator by using distance measures and a reordering of arguments that depends on order-inducing variables. The IOWAD generalizes the OWAD operator and provides a parameterized family of distance aggregation operators between the maximum and the minimum distance. Merigó and Casanovas presented an induced Euclidean ordered weighted averaging distance (IEOWAD) operator, which uses the IOWA operator and the Euclidean distance in the same formulation. Going a step further, Merigó and Casanovas introduced the induced generalized OWA distance (IGOWAD) (or induced Minkowski OWA distance (IMOWAD) operator), which generalizes the OWD measure, the OWAD operator, the IOWAD operator, the IEOWAD

* Corresponding author - Tel: +86013920939678; fax: +8602223781295; E-mail: idmake@163.com

operator and a lot of other particular cases. It is very useful for decision-making problems because it can establish a comparison between an ideal, though unrealistic, alternative and available options in order to find the optimal choice. As such, the optimal choice is the alternative closest to the ideal one. The main advantage of the IGOWAD operator is that it is able to deal with complex attitudinal characters (or complex degrees of orness) in the aggregation process. Therefore, we are able to deal with more complex situations more close to the real world.

Usually, when using the IGOWAD operator and above distance measures, it is assumed that the available information is clearly known and can be assessed with exact numbers. However, this may not be the real-life situation found in the decision-making problems because often the available information is vague or imprecise, or it is not possible to analyse the situation with exact numbers. In this case, a better approach may be the use of fuzzy numbers (FNs). With the use of FNs, we are able to analyse the best and worst possible scenarios and the possibility that the internal values of the fuzzy interval will occur. The fuzzy numbers are highly useful in depicting uncertainty and vagueness of an object, and thus can be used as a powerful tool to express data information under various different uncertain environments, which has attracted great attentions. Thus it is necessary to extend the IGOWAD operator and above distance measures to accommodate these situations.

For doing so, in this paper we will develop the fuzzy induced generalized ordered weighted averaging distance (FIGOWAD) operator (or fuzzy induced Minkowski OWA distance (FIMOWAD) operator), which is a generalization of the OWA operator that uses fuzzy numbers, distance measures, order inducing variables and generalized means in order to provide a more general formulation. The FIGOWAD includes a wide range of distance operators such as the fuzzy maximum distance, the fuzzy minimum distance, the fuzzy normalized generalized distance (FNGD), the fuzzy weighted generalized distance (FWGD), the fuzzy generalized ordered weighted averaging distance (FGOWAD) operator, the fuzzy induced ordered weighted averaging distance (FIOWAD) operator and the fuzzy induced Euclidean ordered weighted averaging distance (FIEOWAD) operator. We study some families of the FIGOWAD operators. The main advantage of the FIGOWAD is that it is able to deal with complex reordering processes that represent a wide range of factors in an uncertain environment that can be assessed with fuzzy numbers. Then, the decision-making problem can be represented more completely because we now consider the best and worst possible scenarios. Another advantage is that it is able to deal with complex attitudinal characters (or complex degrees of orness) in the decision process by using order-inducing variables. Finally, develop an application of the new operator in a

strategic group decision-making problem under fuzzy environment.

This paper is organized as follows. Section 2 presents some basic concepts. In Sect. 3, we present the FIGOWAD operator and Sect. 4 we develop an application in decision making. Finally, Sect. 5 summarizes the main conclusions of the paper.

2 Preliminaries

The distance measures are very useful techniques that have been used in a wide range of applications such as fuzzy set theory, decision-making, operational research, etc. The generalized (or Minkowski) distance is one of the most widely used distance measures which generalizes a wide range of other distances such as the Hamming distance, the Euclidean distance, etc. For two sets, $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$, they can be described as follows.

Definition 1. A normalized generalized distance (NGD) of dimension n is a mapping $NGD: R^n \rightarrow R$, which has the following form:

$$NGD(A, B) = \left(\frac{1}{n} \sum_{i=1}^n |a_i - b_i|^\lambda \right)^{1/\lambda}, \quad (1)$$

where a_i and b_i is the i^{th} arguments of the sets A and B and λ is a parameter such that $\lambda \in (-\infty, +\infty)$. If we give different values to the parameter λ , we can obtain a wide range of special cases. For example, if $\lambda = 1$, we obtain the normalized Hamming distance. If $\lambda = 2$, the normalized Euclidean distance.

Sometimes, when normalizing the generalized distance, we prefer to give different weights to each individual distance. In this case, the distances are known as the weighted generalized distance, which can be defined as follows, respectively:

Definition 2. A weighted generalized distance (WGD) of dimension n is a mapping $WGD: R^n \rightarrow R$ that has an associated weighting $w = (w_1, w_2, \dots, w_n)$ with

$$w_j \in [0, 1] \text{ and } \sum_{j=1}^n w_j = 1 \text{ such that:}$$

$$WGD = \left(\sum_{i=1}^n w_i |a_i - b_i|^\lambda \right)^{1/\lambda}, \quad (2)$$

where a_i and b_i is the i^{th} arguments of the sets A and B and λ is a parameter such that $\lambda \in (-\infty, +\infty)$. If $\lambda = 1$, we obtain the weighted Hamming distance (WHD). If $\lambda = 2$, the weighted Euclidean distance (WED).

The IOWA operator is an extension of the OWA operator. The main difference is that the reordering step

is not carried out with the values of the argument a_i . In this case, the reordering step is developed with order-inducing variables that reflect a more complex reordering process. The IOWA operator also includes as particular cases maximum, minimum and average criteria. It can be defined as follows:

Definition 3. An IOWA operator of dimension n is a mapping IOWA: $R^n \times R^n \rightarrow R$ that has an associated weighting W with $w_j \in [0,1]$ and $\sum_{j=1}^n w_j = 1$ such that:

$$IOWA(\langle u_1, a_1 \rangle, \dots, \langle u_n, a_n \rangle) = \sum_{j=1}^n w_j b_j, \tag{3}$$

where b_j is a_i value of the IOWA pair $\langle u_i, a_i \rangle$ having the j^{th} largest u_i , u_i is the order inducing variable and a_i is the argument variable.

The IGOWAD (or IMOWAD) operator is a distance measure that uses the IOWA operator in the normalization process of the Minkowski distance. Then, the reordering of the individual distances is developed with order inducing variables. For two sets $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$, the IGOWAD operator can be defined as follows:

Definition 4. An IGOWAD operator of dimension n is a mapping $f : R^n \times R^n \times R^n \rightarrow R$ that has an associated weighting W with $w_j \in [0,1]$ and $\sum_{j=1}^n w_j = 1$ such that:

$$f(\langle u_1, a_1, b_1 \rangle, \dots, \langle u_n, a_n, b_n \rangle) = \left(\sum_{j=1}^n w_j d_j^\lambda \right)^{1/\lambda}, \tag{4}$$

where d_j is the $|a_i - b_i|$ value of the IGOWAD triplet $\langle u_i, a_i, b_i \rangle$ having the j largest u_i , u_i is the order inducing variable, $|a_i - b_i|$ is the argument variable represented in the form of individual distances and λ is a parameter such that $\lambda \in (-\infty, +\infty)$. Especially, if $\lambda = 1$, then the IGOWAD is called the induced ordered weighted averaging distance (IOWAD) operator [30], and if $\lambda = 2$, then the induced Euclidean ordered weighted averaging distance (IEOWAD) operator [32].

When using the IGOWAD operator, it is assumed, that the available information is represented in the form of exact numbers. However, this may not be the real situation found in the decision-making problem. Sometimes the available information is vague or imprecise and it is not possible to analyse it with exact numbers. In this case, it is more suitable to use linguistic variables to assess the uncertainty. In the following, we shall develop the linguistic induced generalized ordered weighted averaging distance (LIGOWAD) operator.

3 Fuzzy induced generalized ordered weighted averaging distance (LIGOWAD) operator

The fuzzy numbers are highly useful in depicting uncertainty and vagueness of an object, and thus can be used as a powerful tool to express data information under various different fuzzy environments, which has attracted great attentions there are many papers concerning the fuzzy number arithmetic. For practical reasons we use, however, the notation introduced by Van Laarhoven and Pedrycz [44]. According to this notation, a triangular fuzzy number \hat{a} may be expressed as following:

Definition 5. Let $\hat{a} = [a^L, a^M, a^U]$, where $a^L \leq a^M \leq a^U$, then \hat{a} is called a triangular fuzzy number (TFN), where a^L and a^U stand for the lower and upper values of \hat{a} , and a^M stands for the modal value. Especially, if $a^L = a^M = a^U$, then \hat{a} is reduced to a real number.

Let $\hat{a} = [a^L, a^M, a^U]$ and $\hat{b} = [b^L, b^M, b^U]$ be two triangular fuzzy numbers, below we first introduce some operational laws of triangular fuzzy numbers as follows:

- (1) $\hat{a} + \hat{b} = [a^L + b^L, a^M + b^M, a^U + b^U]$;
- (2) $\lambda \hat{a} = [\lambda a^L, \lambda a^M, \lambda a^U]$, where $\lambda \geq 0$. Especially, $\hat{a} \cdot 0 = 0$, if $\hat{a} = 0$.

Definition 6. Let $\hat{a} = [a^L, a^M, a^U]$ and $\hat{b} = [b^L, b^M, b^U]$ be two triangular fuzzy numbers, then:

$$d(\hat{a}, \hat{b}) = \sqrt{\frac{1}{3} (a^L - b^L)^2 + (a^M - b^M)^2 + (a^U - b^U)^2}, \tag{5}$$

is called a distance between \hat{a} and \hat{b} .

The fuzzy induced generalized ordered weighted averaging distance (FIGOWAD) operator is an extension of the IGOWAD operator that uses uncertain information in the aggregation represented in the form of fuzzy numbers. The reason for using this operator is that sometimes, the uncertain factors that affect our decisions are not clearly known and in order to assess the problem we need to use fuzzy numbers in order to consider the different uncertain results that could happen in the future. Note that the FIGOWAD operator can also be seen as an aggregation operator that uses the main characteristics of the IOWA, distance measures and fuzzy numbers. Moreover, it also uses a complex reordering process by using order inducing variables. Let Ψ be the set of all triangular fuzzy numbers, for two collections of fuzzy numbers $A = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$ and $B = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_n)$, it can be defined as follows.

Definition 7. A FIGOWAD operator of dimension n is a mapping FIGOWAD: $R^n \times \Psi^n \times \Psi^n \rightarrow R$ that has an

associated weighting W with $w_j \in [0,1]$ and $\sum_{j=1}^n w_j = 1$ such that:

$$FIGOWAD(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle) = \left(\sum_{j=1}^n w_j d_j^\lambda \right)^{1/\lambda}, \quad (6)$$

where d_j is $d(\hat{a}_i, \hat{b}_i)$ value of the FIGOWAD pair $\langle u_i, \hat{a}_i, \hat{b}_i \rangle$ having the j^{th} largest u_i , u_i is the order inducing variable and $d(\hat{a}_i, \hat{b}_i)$ is the argument variable represented in the form of individual distances and λ is a parameter such that $\lambda \in (-\infty, +\infty)$.

Example 1. Let $A = ([3,4,5], [6,7,9], [4,6,7], [2,4,5])$ and $B = ([4,6,7], [3,4,6], [2,5,7], [3,4,6])$ be two collections fuzzy numbers, then

$$d(\hat{a}_1, \hat{b}_1) = \sqrt{\frac{1}{3}(3-4)^2 + (4-6)^2 + (5-7)^2} = 2.16.$$

Similarly, we have $d(\hat{a}_2, \hat{b}_2) = 3$, $d(\hat{a}_3, \hat{b}_3) = 1.29$, $d(\hat{a}_4, \hat{b}_4) = 0.82$.

Assume that both sets have the same order-inducing variables $U = (6,7,3,9)$. Assume the following weighting vector $W = (0.1, 0.2, 0.2, 0.5)$ and without loss of generality, let $\lambda = 2$, then we can calculate the distance between A and B by using the FIGOWAD operator: $FIGOWAD(A, B) = 1.98$.

From a generalized perspective of the reordering step, we can distinguish between the descending LIGOWAD (DLIGOWAD) operator and the ascending LIGOWAD (ALIGOWAD) operator by using $w_j = w_{n-j+1}^*$, where w_j is the j^{th} weight of the DLIGOWAD and $w_j = w_{n-j+1}^*$ is the j^{th} weight of the ALIGOWAD operator.

Note that if the weighting vector is not normalized, i.e., $W = \sum_{j=1}^n w_j \neq 1$, then, the LIGOWAD operator can be expressed as:

$$FIGOWAD(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle) = \left(\frac{1}{W} \sum_{j=1}^n w_j d_j^\lambda \right)^{1/\lambda}. \quad (7)$$

Similar to the IGOWAD operator, the FIGOWAD operator is commutative, monotonic, bounded and idempotent. Another interesting issue is the problem of ties in the order inducing variables. As it was explained by Yager and Filev, the easiest way to solve this problem consists in replacing each argument of the tied inducing variables by its linguistic normalized linguistic generalized distance.

Analysing the applicability of the FIGOWAD operator, we can see that it is applicable to similar situations already discussed in other types of induced aggregation operators where it is possible to use linguistic information. For example, we could use it in different decision making problems, etc.

The FIGOWAD operator provides a parameterized family of aggregation operators. Basically, we distinguish between the families found in the weighting vector W and those found in the parameter λ .

If we analyse the parameter λ , we can find a wide range of distance measures such as the FLOWAD, the FIEWAD, the fuzzy induced ordered weighted geometric distance (FIOWGD) operator, the fuzzy induced ordered weighted harmonic averaging distance (FIOWHAD) operator and a lot of other cases.

Remark 1. If $\lambda = 1$, then, we get the FLOWAD operator.

$$FIGOWAD(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle) = \sum_{j=1}^n w_j d_j^\lambda. \quad (8)$$

Note that if $w_j = 1/n$ for all j , we get the fuzzy normalized Hamming distance (FNHD). The fuzzy weighted Hamming distance (FWHD) is obtained if $u_i > u_{i+1}$ for all i , and the fuzzy ordered weighted averaging distance (FOWAD) is obtained if the ordered position of u_i is the same as the ordered position of d_j such that d_j is the j^{th} largest of $d(\hat{a}_i, \hat{b}_i)$.

Remark 2. If $\lambda = 2$, then we get the FIEWAD operator.

$$FIGOWAD(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle) = \left(\sum_{j=1}^n w_j d_j^2 \right)^{1/2}. \quad (9)$$

Note that if $w_j = 1/n$ for all j , we get the fuzzy normalized Euclidean distance (FNED). The fuzzy weighted Euclidean distance (FWED) is obtained if $u_i > u_{i+1}$ for all i , and the fuzzy Euclidean ordered weighted averaging distance (FEOWAD) is obtained if the ordered position of u_i is the same as the ordered position of d_j such that d_j is the j^{th} largest of $d(\hat{a}_i, \hat{b}_i)$.

Remark 3. When $\lambda = 0$, we get the FLOWGD operator.

$$FIGOWAD(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle) = \prod_{j=1}^n d_j^{w_j}. \quad (10)$$

Remark 4. When $\lambda = -1$, we get the FLOWHAD operator.

$$FIGOWAD\left(\langle u_1, \hat{a}_1, \hat{b}_1 \rangle, \dots, \langle u_n, \hat{a}_n, \hat{b}_n \rangle\right) = \frac{1}{\sum_{j=1}^n \frac{w_j}{d_j}} \quad (11)$$

By choosing a different manifestation of the weighting vector in the FIGOWAD operator, we are able to obtain different types of distance aggregation operators. For example, we can obtain the fuzzy maximum distance, the fuzzy minimum distance, the FNGD, the FWGD, the FGOWAD, the Step-FIGOWAD and the Olympic-FIGOWAD

If $w_j = 1/n$, we get the FNGD.

The fuzzy maximum distance is obtained if $w_p = 1$, $w_j = 0$, for all $j \neq p$, and $u_p = \text{Max}\{d(\hat{a}_i, \hat{b}_i)\}$.

The fuzzy minimum distance is obtained if $w_p = 1$, $w_j = 0$, for all $j \neq p$, and $u_p = \text{Min}\{d(\hat{a}_i, \hat{b}_i)\}$.

The FWGD is obtained if $u_i > u_{i+1}$ for all i .

The FGOWAD operator is obtained if the ordered position of u_i is the same as the ordered position of d_j such that d_j is the j^{th} largest of $d(\hat{a}_i, \hat{b}_i)$.

Step-FIGOWA: If $w_k = 1$ and $w_j = 0$ for all $j \neq k$.

Olympic-FIGOWAD: If $w_1 = w_n = 0$ and for all others $w_j = 1/(n-2)$.

Remark 5. Using a similar methodology, we could develop numerous other families of FIGOWAD operators. For more information, refer to [30-32].

4 Illustrative Example

The FIGOWAD operator can be applied in a wide range of problems such as statistics, engineering, economics, decision theory and clustering under fuzzy environment. In this paper, we will consider a decision-making application in the selection of strategies by using a group analysis. The process to follow in the selection of investments with the FIGOWAD operator in group decision-making is similar to the process developed in Ref. [32], [33], with the difference that now we are considering an uncertain situation where the group of experts of the company needs to assess the available information with fuzzy numbers.

Assume a company that operates in North America and Europe is analysing the general policy for the next year and they consider five possible strategies to follow (adopted from Ref. [35]):

A_1 is a computer company. A_2 is a chemical company. A_3 is a food company. A_4 is a car company. A_5 is a TV company.

In order to evaluate these strategies, the group of experts considers that the key factor is the economic situation of the next year. Thus, depending on the

situation, the expected benefits will be different. The experts have considered five possible situations for the next year:

S_1 = Negative-growth rate. S_2 = Growth rate near 0.

S_3 = Low-growth rate. S_4 = Medium-growth rate. S_5 = High-growth rate.

The group of experts of the company is constituted by three persons that give its own opinion about the uncertain expected results that may occur in the future. The expected results depending on the situation S_i and the alternative A_k are shown in Tables 1-3. Note that the results are TFNs.

TABLE 1 Fuzzy Payoff Matrix-expert 1

	S_1	S_2	S_3	S_4	S_5
A_1	[60,70,80]	[30,40,50]	[50,60,70]	[70,80,90]	[30,40,50]
A_2	[50,60,70]	[70,80,90]	[20,30,40]	[50,60,70]	[40,50,60]
A_3	[10,20,30]	[30,40,50]	[40,50,60]	[60,70,80]	[70,80,90]
A_4	[20,30,40]	[40,50,60]	[60,70,80]	[80,90,100]	[70,80,90]
A_5	[30,40,50]	[40,50,60]	[70,80,90]	[20,30,40]	[60,70,80]

TABLE 2 Fuzzy Payoff Matrix-expert 2

	S_1	S_2	S_3	S_4	S_5
A_1	[50,60,70]	[70,80,90]	[80,90,100]	[20,30,40]	[70,80,90]
A_2	[60,70,80]	[20,30,40]	[50,60,70]	[30,40,50]	[40,50,60]
A_3	[60,70,80]	[50,60,70]	[20,30,40]	[70,80,90]	[60,70,80]
A_4	[70,80,90]	[10,20,30]	[40,50,60]	[70,80,90]	[20,30,40]
A_5	[30,40,50]	[50,60,70]	[40,50,60]	[50,60,70]	[30,40,50]

TABLE 3 Fuzzy Payoff Matrix-expert 3

	S_1	S_2	S_3	S_4	S_5
A_1	[20,30,40]	[60,70,80]	[40,50,60]	[70,80,90]	[40,50,60]
A_2	[40,50,60]	[10,20,30]	[70,80,90]	[50,60,70]	[50,60,70]
A_3	[40,50,60]	[70,80,90]	[80,90,100]	[30,40,50]	[50,60,70]
A_4	[70,80,90]	[80,90,100]	[20,30,40]	[50,60,70]	[60,70,80]
A_5	[60,70,80]	[50,60,70]	[60,70,80]	[30,40,50]	[60,70,80]

According to their objectives, the company establishes the following collective ideal investment shown in Table 4.

TABLE 4 Ideal Strategy

	S_1	S_2	S_3	S_4	S_5
I	[80,90,100]	[80,90,100]	[80,90,100]	[80,90,100]	[80,90,100]

In order to aggregate the information, the group of experts calculates the attitudinal character of the candidate. Due to the fact that the attitudinal character depends upon the opinion of several members of the board of directors, it is very complex. Therefore, they

need to use order-inducing variables in the reordering process. The results are shown in Table 5.

TABLE 5 Order-inducing Variables

	S_1	S_2	S_3	S_4	S_5
A_1	10	8	15	17	24
A_2	15	12	9	18	20
A_3	16	14	12	10	8
A_4	13	22	17	15	9
A_5	20	25	18	14	16

With this information, we can make an aggregation to make a decision. First, we aggregate the information of the three experts to obtain a unified payoff matrix. We use the FWA operator to obtain this matrix while assuming that $V = (0.3, 0.3, 0.4)$. The results are shown in Table 6.

TABLE 6 Collective Fuzzy Payoff Matrix

	S_1	S_2	S_3	S_4	S_5
A_1	[41,51,61]	[54,64,74]	[55,65,75]	[55,65,75]	[46,56,76]
A_2	[49,59,69]	[31,41,51]	[49,59,69]	[44,54,64]	[44,54,64]
A_3	[37,47,57]	[52,62,72]	[50,60,70]	[51,61,71]	[59,69,79]
A_4	[55,65,75]	[47,57,67]	[38,48,58]	[65,75,85]	[51,61,71]
A_5	[42,52,62]	[47,57,67]	[57,67,77]	[33,43,54]	[51,61,71]

With this information, it is possible to aggregate the available information in order to take a decision. The method consists in comparing the available investments with the ideal one by using the FIGOWAD operator and its particular cases. The optimal choice would be the alternative closest to the ideal. In this example, we consider the fuzzy maximum distance, the fuzzy minimum distance, the FNHD, the FWHD, the FOWAD, the FLOWAD, the FWED, the FLOWAD and the FIEWAD operators. We assume the following weighting vector $W = (0.1, 0.2, 0.2, 0.2, 0.3)$. The results are shown in Table 7.

TABLE 7 Aggregated Results

	FND	FWHD	FOWAD	FLOWAD	FWED	FLOWAD	FIEWAD
A_1	0.43	0.26	0.24	0.26	0.41	0.19	0.28
A_2	0.34	0.24	0.23	0.24	0.29	0.22	0.23
A_3	0.56	0.17	0.27	0.20	0.21	0.15	0.16
A_4	0.42	0.21	0.20	0.22	0.33	0.16	0.21
A_5	0.35	0.25	0.22	0.23	0.26	0.18	0.25

As we can see, depending on the aggregation operator used, the ordering of the strategies may be different. Therefore, the decision about which strategy select may be also different. If we establish an ordering of the

investments, a typical situation if we want to consider more than one alternative, we will get the following orders shown in Table 8. Note that the first alternative in each ordering is the optimal choice.

TABLE 8 Ordering of the Strategies

	Ordering
FND	$A_2 \succ A_4 \succ A_5 \succ A_1 \succ A_3$
FWHD	$A_3 \succ A_4 \succ A_2 \succ A_5 \succ A_1$
FOWAD	$A_4 \succ A_5 \succ A_2 \succ A_1 \succ A_3$
FLOWAD	$A_3 \succ A_4 \succ A_5 \succ A_1 \succ A_2$
FWED	$A_3 \succ A_5 \succ A_2 \succ A_4 \succ A_1$
FLOWAD	$A_3 \succ A_4 \succ A_5 \succ A_2 \succ A_1$
FIEWAD	$A_3 \succ A_4 \succ A_2 \succ A_5 \succ A_1$

5 Conclusions

We have presented a wide range of fuzzy induced generalized aggregation distance operators. First, we have introduced the FIGOWAD operator. It is a generalization of the OWA operator that uses order-inducing variables in order to assess complex reordering processes, distance measure, fuzzy information and generalized means. We have analysed some of its main properties. We have seen that it generalizes a wide range of distance aggregation operators such as the FNGD, the FGOWAD and the FOWAD operator.

We have also developed an application of the new approach in a strategic decision-making problem. We have seen that the FIGOWAD is very useful because it represents very well the uncertain information by using IFNs. We have also seen that depending on the particular case of the FIGOWAD operator used the results may lead to different decisions.

In future research, we expect to develop further improvements by adding more characteristics in the model such as the use of other types of aggregation operators and apply it in other decision-making problems.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China(No.71271191,70871111), Statistical Scientific Key Research Project of China (No. 201LZ48), the Key Research Centre of Philosophy and Social Science of Zhejiang Province – Modern Port Service Industry and Creative Culture Research Centre, Ministry of Education Humanities and Social Sciences Fund (11YJC790141), Zhejiang Provincial Key Research Base for Humanities and Social Science Research (Statistics), Projects in Science and Technique of Ningbo Municipal (No. 2012B82003), the MOE Project of Key Research Institute of Humanities and Social Sciences in Universities (No. 13JJD910002), and the Key Research Project of Zhejiang Wanli University.

References

- [1] Beliakov G, Pradera A, Calvo T 2007 *Aggregation functions: A guide for practitioners* Press: Springer Publishing Company
- [2] Calvo T, Mayor G, Mesiar R 2002 *Aggregation operators: new trends and applications* 97 Springer
- [3] Xu Z S, Da Q L 2003 An overview of operators for aggregating information *International Journal of Intelligent Systems* 18(9) 953-69
- [4] Yager R R 1988 On ordered weighted averaging aggregation operators in multi-criteria decision making *IEEE Transactions on Systems, Man and Cybernetics B* 18(1) 183-90
- [5] Ahn B S 2009 Some remarks on the LSOWA approach for obtaining OWA operator weights *International Journal of Intelligent Systems* 24(12) 1265-79
- [6] Amin G R, Emrouznejad A 2006 An extended minimax disparity to determine the OWA operator weights *Computers & Industrial Engineering* 50(3) 312-16
- [7] Cheng C H, Wang J W, Wu M C 2009 OWA-weighted based clustering method for classification problem *Expert Systems with Applications* 36(3) 4988-95
- [8] Filev D P, Yager R R 1998 On the issue of obtaining OWA operator weights *Fuzzy Sets and Systems* 94(2) 157-69
- [9] Herrera F, Herrera-Viedma E 1997 Aggregation operators for linguistic weighted information *IEEE Transactions on Systems, Man and Cybernetics A* 27(5) 646-55
- [10] Kacprzyk J, Zadrozny S 2009 Towards a generalized and unified characterization of individual and collective choice functions under fuzzy and nonfuzzy preferences and majority via ordered weighted average operators *International Journal of Intelligent Systems* 24(1) 4-26
- [11] Karayiannis N 2000 Soft learning vector quantization and clustering algorithms based on ordered weighted aggregation operators *IEEE Transactions on Neural Networks* 11(5) 1093-105
- [12] Liu J W, Cheng C H, Chen Y H, Chen T L 2010 OWA Rough set model for forecasting the revenues growth rate of the electronic industry. *Expert Systems with Applications* 37(1) 610-17
- [13] Liu X 2008 A general model of parameterized OWA aggregation with given orness level *International Journal of Approximate Reasoning* 48(2) 598-627
- [14] Merigó J M 2010 Fuzzy decision making with immediate probabilities *Computers & Industrial Engineering* 58(4) 651-57
- [15] Merigó J M, Casanovas M 2010 The fuzzy generalized OWA operator and its application in strategic decision making *Cybernetics & Systems* 41(5) 359-70
- [16] Merigó J M, Casanovas M 2010 Induced and Heavy aggregation operators with distance measures *Journal of Systems Engineering and Electronics* 21(3) 431-39
- [17] Merigó J M, Wei G W 2011 Probabilistic aggregation operators and their application in uncertain multi-person decision making *Technological and Economic Development of Economy* 17(2) 335-51
- [18] Wang S Q, Li D F, Wu Z Q 2009 Generalized ordered weighted averaging operators based methods for MADM in intuitionistic Fuzzy setting *Journal of Systems Engineering and Electronics* 20(6) 1247-54
- [19] Xu Z S 2004 EOWA and EOWG operators for aggregating linguistic labels based on linguistic preference relations *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12(6) 791-810
- [20] Xu Z S 2005 An overview of methods for determining OWA weights *International Journal of Intelligent Systems* 20(8) 843-65
- [21] Xu Z S 2008 Hybrid weighted distance measures and their application in pattern recognition *Lecture Notes in Computer Science* 17-23
- [22] Yager R R 1993 Families of OWA operators *Fuzzy Sets and Systems* 59(2) 125-48
- [23] Yager R R 2010 Norms Induced from OWA Operators *IEEE Transactions on Fuzzy Systems* 18(1) 57-66
- [24] Zeng S Z, Su W H 2011 Intuitionistic fuzzy ordered weighted distance operator *Knowledge-Based Systems* 24(8) 1224-1232
- [25] Yager R R 2003 Induced aggregation operators *Fuzzy Sets and Systems* 137(1) 59-69
- [26] Yager R R, Feliv D P 1999 Induced ordered weighted averaging operators *IEEE Transactions on Systems, Man and Cybernetics, Part B* 29(2) 141-50
- [27] Chiclana F, Herrera-Viedma E, Herrera F, Alonso S 2007 Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations *European Journal of Operational Research* 182(1) 383-99
- [28] Merigó J M, Casanovas M 2009 Induced aggregation operators in decision making with the Dempster-Shafer Belief structure *International Journal of Intelligent Systems* 24(8) 934-54
- [29] Merigó J M, Casanovas M 2010 Decision making with distance measures and linguistic aggregation operators *International Journal of Fuzzy Systems* 12(3) 190-98
- [30] Merigó J M, Casanovas M 2010 Decision making with distance measures and induced aggregation operators *Computers & Industrial Engineering* 60(1) 66-76
- [31] Merigó J M and Casanovas M 2011 Induced aggregation operators in the Euclidean distance and its application in financial decision making *Expert Systems with Applications* 38(6) 7603-08
- [32] Merigó J M, Casanovas M 2011 A new Minkowski distance based on induced aggregation operators *International Journal of Computational Intelligence Systems* 4(2) 123-33
- [33] Merigó J M, Casanovas M, Zhou L G, Chen H Y 2011 Induced and linguistic generalized aggregation operators and their application in linguistic group decision making *Group Decision and Negotiation* 21(4) 531-49
- [34] Merigó J M, Gil-Lafuente A M 2009 The induced generalized OWA operator *Information Sciences* 179(6) 729-41
- [35] Wei G W 2010 Some induced geometric aggregation operators with intuitionistic fuzzy information and their application to group decision making *Applied Soft Computing* 10(2) 423-31
- [36] Xu Z S 2006 Induced uncertain linguistic OWA operators applied to group decision making *Information Fusion* 7(2) 231-38
- [37] Yager R R 2004 Choquet aggregation using order inducing variables *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12(1) 69-88
- [38] Zeng Z S, Li W, Merigó J M 2013 Extended induced ordered weighted averaging distance operators and their application to group decision-making *International Journal of Information Technology & Decision Making* 12(4) 1-23
- [39] Xu Z S, Chen J 2008 Ordered weighted distance measures *Journal of Systems Science and Systems Engineering* 17(4) 432-45
- [40] Merigó J M, Gil-Lafuente A M 2010 New decision-making techniques and their application in the selection of financial products. *Information Sciences* 180(11) 2085-94
- [41] Zeng S Z, Su W H 2011 Intuitionistic fuzzy ordered weighted distance operator *Knowledge-Based Systems* 24(8) 1224-32
- [42] Zeng S Z 2013 Some intuitionistic fuzzy weighted distance measures and their application to group decision making *Group Decision and Negotiation* 22(2) 281-98
- [43] Merigó J M, Casanovas M, Martínez L 2010 Linguistic aggregation operators for linguistic decision making based on the Dempster-Shafer Theory of Evidence *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 18(3) 287-304
- [44] Van Laarhoven P J M, Pedrycz W 1983 A fuzzy extension of Saaty's priority theory *Fuzzy Sets and Systems* 11(1) 229-41

Authors	
	<p>Zhihong Ma, born in May, 1975, TianJin City, TianJin Province, P.R. China</p> <p>Current position, grades: Associate Professor of the TianJin agricultural University University studies: University Mathematics Teaching and Application of mathematical research. Scientific interest: Study on mathematical model, data mining. Publications: more than 20 papers published in various journals. Experience: Graduated from Jilin University in 2000, has completed 6 scientific research projects; more than 20 papers published in various journals.</p>
	<p>Jianping Chen, born in December, 1976, Ningbo City, Zhejiang Province, P.R. China</p> <p>Current position, grades: the lecturer of the Ningbo Institute of Technology Zhejiang University. University studies: Service Innovation, Aggregation Operators, Decision Making and Comprehensive Scientific interest: Service Innovation, Aggregation Operators, Decision Making and Comprehensive Publications: more than 6 papers published in various journals. Experience: Jianping Chen has a PhD degree in management from Zhejiang Gongshang University, China. He is a full-time Lecture in Ningbo Institute of Technology Zhejiang University.</p>
	<p>Shouzhen Zeng, born in September, 1981, Ningbo City, Zhejiang Province, P.R. China</p> <p>Current position, grades: the lecturer in Zhejiang Wanli University. University studies: Statistics Research, Knowledge-based Systems and Group decision and Negotiation Scientific interest: Aggregation Operators, Decision Making, Comprehensive evaluation and Uncertainty Publications: 40 papers in journals, books and conference proceedings including journals such as Statistics Research, Experience: He graduated from Zhejiang Gongshang University and obtained the PhD degree in applied statistics in 2013.</p>

Analysis on the effectiveness of China's macroeconomic policy based on the modified Mundell-Fleming model during the post-financial crisis period

Bi Jianxin^{1, 2*}, Lei Lianghai²

¹ Faculty of Management, University of Shanghai for Science and Technology, 516 Jun Gong Road, Shanghai, China

² Faculty of Computer and Information, Zhejiang Wanli University, No.8, South Qian Hu Road Ningbo, Zhejiang, China

Received 1 March 2014, www.tsi.lv

Abstract

This paper proposed the hypotheses of Mundell-Fleming model applicable to current Chinese economic environment, modified the traditional Mundell-Fleming model and analysed the effectiveness of fiscal and monetary policies under different exchange rate systems. Under a fixed exchange rate system, the monetary policy causes economic instability and aggravates economic inequality, thus increasing the difficulty in policy intervention, in which case the fiscal policy has significant results only with the positive coordination of monetary policy; under a floating exchange rate system, the fiscal policy has remarkable effects and the monetary policy has effects which are not uncertain. Finally, the paper analyzes the effectiveness of China's macroeconomic policy using the modified Mundell-Fleming model and proposes the orientation for China's macroeconomic policy in post-crisis period.

Keywords: Mundell-Fleming Model, Sub-prime Mortgage Crisis, Financial Crisis, Fiscal Policy, Monetary Policy

1 Introduction

In 2007, the financial crisis caused by American sub-prime mortgage crisis brought a great impact on Chinese economy through different channels. Many domestic and foreign scholars have made much research on the financial crisis's effects on Chinese economy and Chinese economic policy orientation in post-crisis era. Yiping H [1], Ping L [2] and Sanlin J [3] qualitatively analysed the impact of financial crisis on China's economy in post-crisis period, and proposed Chinese economic policy orientation to deal with the impact of financial crisis in post-crisis period; Rui C [4] analysed the effectiveness of China's fiscal and monetary policy using the Mundell-Fleming model (hereafter referred as M-F model) and proposed the short term, mid-term and long-term adjustment directions for the stable development of Chinese macro-economy under different foreign exchange system levels in the post-crisis period.

In the literatures above, there are some shortcomings in their analyses on the effects of financial crisis on Chinese economy. Most scholars analysed from the qualitative angle, but failed to make scientific judgments on Chinese economic policy orientation and the adjustment policy of macro-economy in post-crisis period. Rui C [4] added the fourth curve of full employment into M-F model based on the traditional M-F model and then analysed the effectiveness of Chinese fiscal & monetary policies. However, it is controversial

about whether the hypothetic conditions based on traditional M-F model are suitable for current situation of China. Boke J [5] believed the hypothesis that "the balance of international payments line (BP line) is a straight line with a positive slope" in M-F model did not conform to the reality of the country with asymmetric capital controls (such as China), so he modified the BP line of traditional M-F model. The modified M-F model hypothesis is under the asymmetric capital control, BP curve shows as a broken line and its slope relates to the control degree of capital inflow. Xiaohui L [6] studied the policy effectiveness and RMB exchange rate policy under FDI flows and assumed the net capital inflows $K(i)$ formed by FDI was a linear function, which was different from the basic hypothesis of M-F model. The author believed that as to the equilibrium of foreign exchange market, M-F model mainly explored short-term capital flows, thus for the short-term capital, the interest rate is the measurement of returns on investment, so for short-term capital, interest rate is the measurement on the return on investment; while for FDI policy under the modified M-F model; the fourth part analyzes the orientation of China's macroeconomic policy in post-crisis period under different exchange rate systems; the third part is the analysis on the effectiveness of China's macroeconomic. This paper is structured as follows: the second part refers to the modification of M-F model and the analysis on the effectiveness of the fiscal and monetary policies.

* Corresponding author- Tel: +86-150-674-24580 fax: +86-150-674-24580; E-mail: greygirl0511@126.com

Capital utilizing domestic savings, domestic interest rate forms the capital cost of FDI. Jun Wu [7] modified M-F model using the dynamic purchasing power parity theory and examined the modified M-F model.

He pointed out the traditional M-F model's the hypothesis for BP curve didn't conform to Chinese reality; as the modified BP curve had a negative slope, the modified M-F model met Chinese reality. Delei Y [8] et al. believed that in modern economic society, the financial asset price fluctuation has become the most important factor affecting international capital flows in many cases. The price tendency of securities market and real estate market always changed in the direction opposite to that of interest rate, demonstrating the BP curve with a negative slope, thus extending & enriching the form of M-F model. Then they made an empirical test using the economic data of two groups amounting to 28 countries or regions.

This paper believes that the application of any economic theory and model has preconditions suitable for national situation. According to Keynes' foreign trade multiplier theory, the import volume of a country mainly depends on its national income. It means import is the increasing function of national income; while export is mainly decided by foreign income and independent of national income. Therefore, the increase in income may lead to the deterioration of current account. The paper considers the actual economic situation in China currently. Since American financial crisis, the import tariff of China has been decreasing continuously. From the start time of American financial crisis in 2007, Chinese economy in 2008 experienced a great decline comparing with that in 2007. From 2008, Chinese economy entered into a slowdown period, but the current account surplus continued and foreign exchange reserve increased greatly. These phenomenons obviously went against M-F model. Basing on the M-F model modified by the scholars mentioned above and considering that national export depends on not only foreign income but also domestic income, the paper adds the fourth curve of full employment in the modified M-F model to derive eight kinds of internal-external dual-imbalance of macro-economy which are used tools to analyse the internal & external imbalance state of China in post-crisis period. The paper studies the effectiveness of China's macroeconomic policy under the modified M-F model and proposes China's macroeconomics policy orientation in post-crisis period.

2 Modifications of M-F Model and the Analysis on the Effectiveness of Fiscal and Monetary Policies under Different Exchange Rate Systems

2.1 HYPOTHESES OF M-F MODEL UNDER CURRENT CHINA'S ECONOMIC ENVIRONMENT

2.1.1 China is in an incomplete open economic state now

In 2007, American sub-prime mortgage crisis became a global financial crisis soon, seriously impacting developing countries with comparatively low financial development levels. At the same time, the United States put pressure on Chinese RMB appreciation with the excuse of the huge trade surplus of China and China carried out the "double-surplus" open economic policy for a long time. The reason for all of these is trend of global economic integration around the world. As one of developing countries, China is in an incomplete economic situation more or less, so the financial crisis caused by American subprime mortgage crisis influenced Chinese economy to a certain extent.

2.1.2 Under current open economic conditions, the price elasticity is greater than the interest rate plasticity of capital flows in securities market (or real estate market) with capital flows [8]

The traditional M-F model assumes that the net capital outflow is in direct proportion to the differential between international interest rate and domestic interest rate. If not considering the factor of exchange rate, the higher the international interest rate is, the more capitals flow to foreign countries for benefit pursuit, the more the net capitals outflow is, the higher the domestic interest rate is, and the less the capital outflow is. According to the modified M-F model, in a country, when its domestic interest rate falls, its securities market price rises; when its domestic interest rate rises, its securities market price falls. In this case, the fluctuation of financial asset price has become the major factor affecting international capital flows, but interest rate changes are in the direction opposite to the price tendencies in securities market and real estate market [8].

2.1.3 National export depends on not only the foreign income level but also the national income level

According to Mundell, the import volume of a country mainly depends on its national income. To be specific, the import is an increasing function of national income; the export is mainly decided by the foreign in-come and independent of the national income, so income increase may cause the deterioration of current account. The paper believes that Chinese import & export volume has been among the top in the world for a long time. In the more than 30 years since reform and opening, especially the 10 years since entering into WTO, China has realized the best and fastest development. Over the past 10 years, China has become the second biggest economic entity and first export country in the world with a 30% growth in trades and a 10% growth rate of economy. Meanwhile, Chinese GNP was also in the front rank in the world. Under such background, Chinese economic growth contributed to the improvement of production technology and the reduction of production cost per unit, thus

causing the price falling of export commodities and the increase in gross volume of export. Therefore, national export is decided by not only foreign income but also national income.

2.2 DERIVATION OF MODIFIED M-F MODEL

Basing on the hypotheses above, we build models of product market equilibrium, money market equilibrium and external market equilibrium respectively for one country in an open economy.

2.2.1 Product Market Equilibrium Model

In an open economy, the condition for product market equilibrium is $AE=Y$. AE is the effective demand or total willingness expenditure on domestic products and services. It equals to the total expenditure of all domestic sectors, namely $C+I+G$, adds the domestic products and services purchased by foreign people, namely the export, and then subtracts the foreign products and services purchased by domestic people, namely the import M . The equation is as follows:

$$AE = C + I + G + NX = C + I + G + (X - M), \tag{1}$$

where, AE can be expressed by the national income, Y , G means governmental purchase expenditure, I is private investment, $X-M$ is net export, and C is the consumption depending on disposable income Y_d and is expressed as:

$$C = \alpha + \beta Y_d = \alpha + \beta(Y - t_0 + T_r), \tag{2}$$

$$I = e - dr, \tag{3}$$

where t_0 and T_r represent taxation and transfer payment respectively and are the exogenous variables of model. e is autonomous investment. d is the sensitivity of investment to interest rate. r is national interest rate level.

According to traditional M-F model, net export NX equals to export minus input, which depends on the national income and real exchange rate. Then the NX linear function of traditional M-F model is:

$$NX = NX(Y, R) = q - \gamma Y + \frac{nEP_f}{P}, \tag{4}$$

where, Y is national income, Y_w is foreign income exclusive of national income, R is real exchange rate, q is autonomous net export, γ is the marginal propensity to import, EP_f/P is real exchange rate, and n is the net export change ratio caused by real exchange rate change.

The IS curve of traditional M-F model in open economic conditions can be derived from Equations (1), (2), (3) and (4) as:

$$r = \frac{1}{d} \left(\alpha + e + G + q - \beta t_0 + \beta T_r + n \frac{EP_f}{P} \right) - \frac{1}{d} (1 + \gamma - \beta) Y. \tag{5}$$

Considering the real macro-economy environment currently and basing on the third hypothesis above, the net export function can be modified as:

$$NX = NX(Y, Y_w, R) = q - \gamma Y + \frac{nEP_f}{P} + w(Y + Y_w), \tag{6}$$

where $Y, Y_w, R, q, \gamma, EP_f/P$ and n have the same meanings as above and $w (>0)$ is the marginal propensity to export.

The expression of modified IS curve in open economic conditions can be derived from Equation (1), (2), (3) and (5) as:

$$r = \frac{1}{d} \left(\alpha + e + q + G - \beta t_0 + \beta T_r + n \frac{EP_f}{P} + w Y_w \right) - \frac{1}{d} (1 - \beta + \gamma - w) Y. \tag{7}$$

Comparing the absolute values of slope of IS curve before and after modification, the slope after modification is less than the one before modification, indicating the IS curve becomes smoother after modification.

2.2.2 Money Market Equilibrium Model

LM curve represents the combination of interest rate and national income under the equilibrium of money market. M is nominal money supply, and h & k are both parameters representing the sensitivity of money demand to income and interest rate respectively.

$$\frac{M}{P} = M_s = M_d = L_1 + L_2 = kY - hr, \tag{8}$$

$$r = \frac{k}{h} Y - \frac{1}{h} \left(\frac{M}{P} \right). \tag{9}$$

2.2.3 External Market Equilibrium Model

According to the external market equilibrium principles in an open economy, the external equilibrium of a country means the BP account of the country keeps balanced, namely $BP = X - M - F = 0$. In the traditional M-F model, net capital outflow is $F = \sigma(r_w - r)$, where r is domestic interest rate, r_w is the international interest rate, and σ is the parameter representing the sensitivity of capital flows to interest rate. According to the second hypothesis above, the net capital outflow can be expressed as:

$$F = \sigma(R_w - R), \tag{10}$$

where, σ is an parameter, R_w and R are the average rates of return of international and domestic securities markets or real estate markets respectively, and $\frac{\partial F}{\partial R} < 0$. When $R > R_w$, F is negative and represents the net capital inflow. Because the market is not always the same, generally $R_w \neq R$. In general, the domestic interest rate is a decreasing function of average rate of return of securities market or real estate market, namely $\frac{\partial R}{\partial r} < 0$. The functional expression is:

$$R = K - \mu r, \tag{11}$$

where, K and μ are parameters. Generally, $\mu > 1$. Put (11) into (10), we get:

$$F = \sigma(R_w - K + \mu r). \tag{12}$$

According to external market equilibrium principles, $BP = NX - F = 0$. The following equation can be derived from (6) and (12):

$$r = -\frac{1}{\sigma u}(w - \gamma)Y + \frac{1}{\sigma u}\left(q + n\frac{ER_f}{P} - \sigma R_w + \sigma K + wY_w\right). \tag{13}$$

Because $\sigma > 0$ and $u > 1$, $w - \gamma$ depends on the size of marginal propensity to export and marginal propensity to import of China in recent years. Through an empirical study on America and China from 1979 to 2002, Huaimin W [10] concluded that in China the marginal propensity to import was much greater than the marginal propensity to export. Xin Chen [10] drew the same conclusion through his empirical study on America and China from 1994 to 2007. Such is the fact. Chinese economic ranks in the world in recent years indicate the contribution of Chinese economic growth to world economic growth is much greater than the contribution of world economy to Chinese economy. Therefore, $w - \gamma > 0$, namely the slope of BP curve is negative. In the traditional M-F model, BP curve slope γ/σ is positive, while the slope of modified BP curve is negative, which is $-\frac{1}{\sigma u}(w - \gamma)$ with $w - \gamma < \gamma$. It means the BP curve of modified M-F model is smoother than the one of traditional M-F model, and the absolute value of the slope of BP curve of modified M-F model is less than that of traditional M-F model. w and γ are stable in a given period. Therefore, the slope of modified BP curve also depends on the product of capital flow degree (σ) and the sensitivity of domestic securities market or real estate market to interest rate u .

As shown in Figure 1, if a country is in the cross point of curves IS, LM and BP, it reaches a state with the

equilibriums of product market, money market and external market at the same time.

2.3 EIGHT TYPES OF MACROECONOMIC IMBALANCE IN MODIFIED M-F MODEL AFTER ADDING THE FULL EMPLOYMENT

In the open economy, macroeconomic control has two goals: full employment and international payment balance, regarded as internal balance and external balance respectively. To realize internal balance, output level should reach the one of full employment; while to realize external balance, the combination of output and interest rate should be in BP curve. As shown in Figure 1, the ideal case is that the economy is at point F, namely the internal balance of full employment and the external balance of balance of international payment are both achieved at the same time. Away from the equilibrium point, the imbalance, internal imbalance, external imbalance, or internal external double imbalance will be caused. There are two forms of external imbalance, balance of payments surplus and balance of payments deficit; there are also two forms of internal imbalance, economic depression and economic overheating. Considering the single imbalance and double imbalance together, there are eight types of imbalance for a country's macro-economy (as shown in Figure (3-10)).

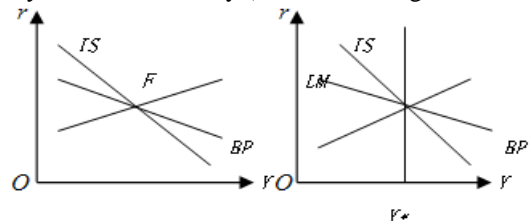


FIGURE 1 Internal-External Double Equilibrium

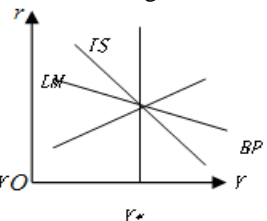


FIGURE 2 Double Equilibrium under Full Employment

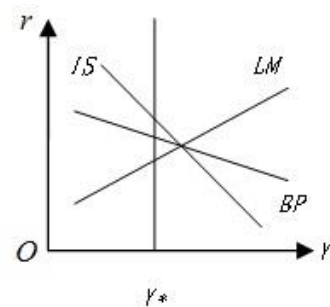


FIGURE 3 Internal Overheating and External Equilibrium

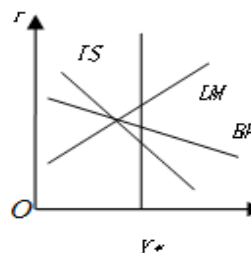


FIGURE 4 Internal Depression and Equilibrium

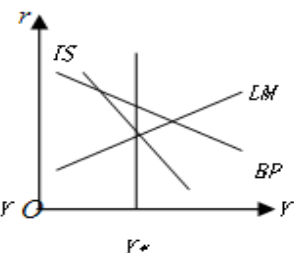


FIGURE 5 Internal Equilibrium and External Surplus

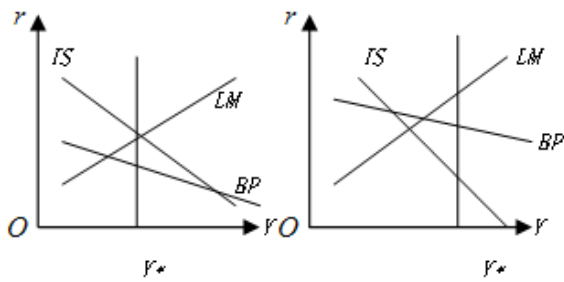


FIGURE 6 Internal Equilibrium and External Deficit

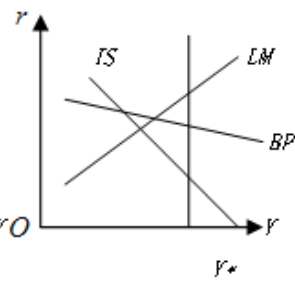


FIGURE 7 Internal and External Surplus

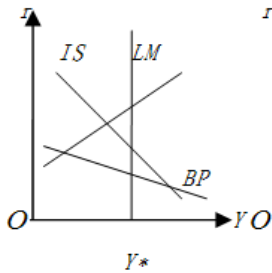


FIGURE 8 Internal Depression and External Deficit

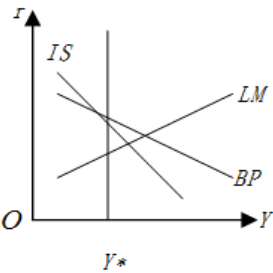


FIGURE 9 Internal Overheating and External Surplus

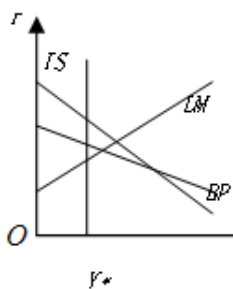


FIGURE 10 Internal Overheating Monetary Policy and External Deficit Rate

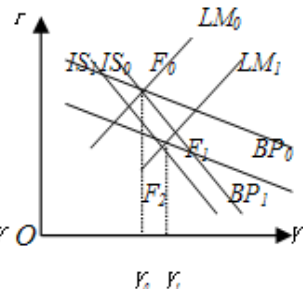


FIGURE 11 Effect of Policy under Floating Exchange System I

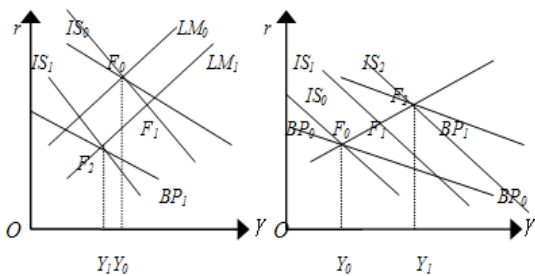


FIGURE 12 Effect of Monetary Policy Floating Exchange Rate System II

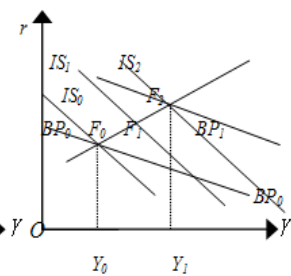


FIGURE 13 Effect of Fiscal Policy under Floating Exchange Rate System

2.4 ANALYSIS ON THE EFFECTIVENESS OF FISCAL AND MONETARY POLICIES UNDER DIFFERENT EXCHANGE RATE SYSTEMS

The M-F model provides an analysis platform for the effectiveness of macroeconomic policy under different exchange rate systems in open economic conditions. Under the floating international exchange rate system in current open economy, countries can choose the degree of floating when establishing the exchange rate system.

Whether a partial-floating or partial-fixed exchange rate system, it will impact the effectiveness of fiscal and monetary policies to different extents.

2.4.1 Fiscal and Monetary Policy Effect under the Floating Exchange Rate System

In the open economy, when a country executes a floating exchange rate system, the exchange rate is decided by the foreign exchange demand curve and foreign exchange supply curve jointly. The demand on foreign exchange is rooted in the import and capital transfer to foreign countries, while the supply is rooted in the export and capital transfer into the country. Therefore, when there is a surplus of external balance, namely $BP > 0$, the supply exceeds the demand in the foreign exchange market and exchange rate decreases which indicates a pressure on currency appreciation. On the contrary, when there is a deficit of external balance, the exchange rate increases which indicates the currency depreciation.

First, we analyse the monetary policy under a floating exchange rate system. Set point F_0 in Figure 11 as the initial equilibrium point. The government implements an expansionary monetary policy. LM_0 moves to LM_1 and F_1 is in the area under BP curve. The balance of international payments surplus leads to the appreciation of domestic currency and thus reduces net export, in which case IS curve shifts to the left from IS_0 to IS_1 and BP curve also shifts to the left from BP_0 to BP_1 , producing a new equilibrium point F_2 . Comparing with point F_0 before government intervention, point F_2 's interest rate decreases and the national income increases. However, if the appreciation of local currency leads to a significant decrease in net export, namely a high elasticity of exchange rate of net export, curves IS and BP may shift to left greatly, and the national income may decrease after the implementation of policy, which is shown by F_2 in Figure 12. We can draw the same conclusion in the case that BP curve is steeper than IS. According to analysis above, under the BP curve with negative slope and the floating exchange system, the government's monetary policy has effects, but such effects are uncertain. Therefore, when adopting the monetary policy, the government should adjust the policy slightly and make further decisions according to practical effects.

Second, we analyse the fiscal policy under a floating exchange rate system. As shown in Figure 13, if the government adopts an expansionary fiscal policy to make IS_0 shift to IS_1 , then the internal equilibrium point moves from F_0 to F_1 . F_1 is above BP_0 curve, in which case the deficit of the balance of international payments and the depreciation of local currency exist to increase the net export, making IS curve move from IS_1 to IS_2 , BP curve move from BP_0 to BP_1 and F_2 become the effective equilibrium point of fiscal policy under floating exchange rate system. It is obvious that under BP curve with negative slope and floating exchange system, the fiscal

policy has the significant effects. We can draw the same conclusion in the case that BP curve is steeper than IS curve.

2.4.2 Fiscal and Monetary Policies under Pegged Exchange Rate System

Under the fixed exchange rate system, the central bank establishes the exchange rate and promises to take measures to stabilize the exchange rate, which is generally achieved by buying and selling foreign exchange. When supply exceeds demand in the foreign exchange market, namely the tendency of increase in exchange rate, the central bank will sell the foreign exchange. Therefore, the nominal exchange rate will stay unchanged. Due to the assumption of fixed price, the real exchange rate stays unchanged, and IS curve & BP curve will not move because of the rise or decline of exchange rate.

We first analyse the monetary policy under the fixed exchange rate system.

Adopting the expansionary monetary policy as shown in Figure 14, LM_0 moves to LM_1 , the interest rate falls, and then the securities market quotation rises and attracts international capital inflow. In this case, the balance of international payments equilibrium point moves from F_0 to F_1 , causing the balance of international payments surplus and the appreciation of local currency. To maintain a fixed exchange rate, if continue to increase money supply, it will aggravate economic disequilibrium. According to the figure, if money supply decreases now, LM will return to LM_0 , thus maintaining the fixed exchange rate and restoring economic equilibrium. But, if money supply decreases, it may cause the increase in interest rate and then restrain the price from rising in the securities market, and thus result in some international capital outflow and even part of domestic funds may flow into the international securities market to seek a higher rate of return. In this case, LM curve may move to LM_2 , forming the balance of international payments deficit which will challenge the fixed exchange rate and government's policies again.

So we can see under the open fixed exchange rate system in the securities market, the effects of monetary policy always place the government into a dilemmatic situation. It is also the challenge faced by many emerging market economy countries after opening the securities market. Similarly, the tight monetary policy will lead to the same policy effect under the fixed exchange rate system, aggravating the instable and imbalanced situation of economy and increasing the difficulty of policy intervention.

Next is the analysis on the fiscal policy under the fixed exchange rate system.

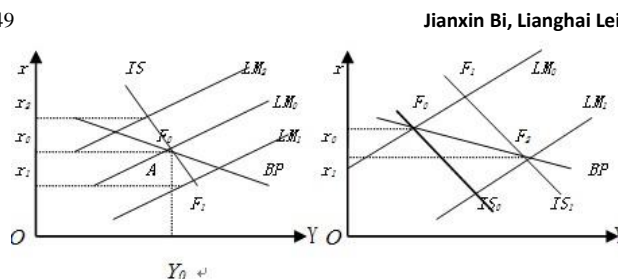


FIGURE 14 Effect of Monetary Policy Fixed Interest Rate

FIGURE 15 Effect of Fiscal under Policy under Fixed Exchange Rate

Adopting the expansionary fiscal policy as shown in Figure 15, IS_0 moves to IS_1 and the internal equilibrium point is F_1 . The increase in income leads to the aggravation of current account, when interest rate rises and obstructs the securities market quotation causing international capital outflow, the aggravation of capital account and the balance of international payments deficit. According to the figure, to maintain the fixed exchange rate, the money supply should be enlarged to make LM_0 move to LM_1 . The three curves intersect at point F_2 , realizing the internal and external economic equilibriums at the same time. Due to monetary ease and interest rate decline, the environment of securities market is improved and then the capital account is improved, making the balance of international payments get better and even return to equilibrium. It indicates that, when the securities market becomes the major channels for the inflows and out-flows of international capital, the fiscal policy has significant effects with the positive coordination of monetary policy.

3 Analyses on the Effectiveness of Chinese Macroeconomic Policy with the Modified M-F Model

3.1 ANALYSIS ON INTERNAL AND EXTERNAL IMBALANCE OF CHINESE MACROECONOMIC SITUATION

The cyclical changes of Chinese economy have objective laws, which is also the inevitable result of the massive accumulation of contradictions during rapid economic growth. In addition, the financial crisis initiated by American sub-prime mortgage crisis further exposed the contradictions aggravated these imbalances among which the internal external economic imbalance is the most obvious.

Chinese economy has kept a double-figure growth in the five years from 2003 to 2007 and also has accumulated a lot of contradictions and problems, showing as the following four major imbalances.

First, economic growth excessively relied on export and the domestic demand is insufficient. The second imbalance is between the overheated investment and the inadequate consumption. The proportion of final consumption, especially household consumption, was too low. Third, the pressure of resource and environment are bigger and bigger. Resource and environment have

become the major "bottle-neck" for the sustainable growth of Chinese economy. Fourth, Chinese capital account and current account began to have double surplus since 1994. The great increase in foreign exchange reserves has brought pressure on the appreciation of RMB till now. The foreign trade surplus of China stayed at a high level from 2007 to 2010 after the sub-prime mortgage crisis. Such situation is different from those of general developing countries obviously. The high domestic savings led to the high domestic investment, and the high investment brought excess capacity which would find a way out abroad when facing the deficiency of domestic market demand, finally causing the expansion of general trade surplus [11].

Among the four imbalances, the imbalance of internal and external demands, the imbalance of investment and consumption, and the imbalance of long-term and continuous double surplus and high foreign exchange reserves are particularly significant, directly constraining the implementation of sustainable economic development strategy in China.

From the angle of unemployment rate, according to the data from the National Bureau of Statistics of China, the unemployment rate remained in the range of 3.6-4.3% from 2001 to 2010. It is the registered urban unemployment rate exclusive of the surplus labour in rural areas and the unregistered unemployment in urban areas. Therefore, the data cannot reflect the unemployment status in China accurately. If including these data, Chinese unemployment rate would be much higher than 4%.

In particular, affected by American financial crisis in 2007, Chinese export-oriented enterprises' ability to absorb employees declined significantly. Industrial restructuring and upgrading also affected small and medium-sized labour-intensive enterprises with a large employment capacity. The backward development of tertiary industry was weak in pushing employment, thus causing a severe employment status in China. According to the Blue Paper issued by the Chinese Academy of Social Sciences on December 16, 2008, the urban unemployment rate in China has rose to 9.4% and reached 14.2% in 2009.

According to the comprehensive analysis on Chinese macro-economy, from 2001 to 2003, Chinese economy experienced the external balance of international payments surplus and internal depression (as shown in Figure 7); from 2004 to 2006, Chinese economy experienced the external balance of international payments surplus and kept a growth state with full employment (as shown in Figure 5); in 2007, Chinese economy experienced the external surplus and its economic growth rate rose to a level higher than the growth of full employment (as shown in Figure 9); from 2008 to 2012, Chinese economy experienced external surplus and its economic growth declined to a level lower than the growth of full employment (as shown in Figure 7).

3.2 IS-LM-BP MODEL FOR CURRENT CHINA

3.2.1 Current Exchange Rate System of China

After the market-oriented reform of exchange rate system in 1994, China began to implement a single and supervisory floating exchange rate system based on market supply & demand. Due to the outbreak of Asian financial crisis in 1997, China declared that RMB would not depreciated against foreign currencies and made RMB tightly pegged to US dollar to fix the exchange. At this time, the International Monetary Fund (IMF) included Chinese exchange rate system into those to be strictly pegged. In July 2005, Chinese monetary authority adjusted the nominal anchor of exchange rate from pegging to US dollar to referring to currency basket, and implemented a managed floating exchange rate system. In July 2008, RMB began bilateral fluctuations and was trans-formed into the floating exchange rate system gradually. Therefore, it can be considered that the system implemented by China currently is a half-floating exchange rate system with some characteristic of partial fixed exchange rate system.

3.2.2 Slope Characteristics of Curves BP, IS and LM in the Post-crisis Period

Before 2005, China adopted a fixed exchange rate system that RMB was pegged to US dollar. Thereafter, RMB exchange rate system began to change, but the government still applied intensive intervention to the foreign exchange market and the annual fluctuation of RMB-dollar exchange rate is extremely small. It is regarded that the RMB exchange rate is actually a fixed exchange rate system "softly pegging to US dollar". Consequently, BP curve of international payment balance basically remains unchanged.

In the first half of 2005, the pilot reform of the foreign exchange administration of overseas investment was promoted to the whole country, greatly increasing the foreign exchange quota for overseas investment and further delegating the power of approval. On July 1, 2006, the State Administration of Foreign Exchange made great adjustments to the foreign exchange management of overseas investment. All of these show that the international payment balance curve (BP curve) of China will become smoother and smoother in the future. The analysis on modified model has shown that the modified BP curve is smoother than the traditional one.

The absolute value of unmodified IS curve's slope is $(1-\beta+\gamma)/d$, where β is the marginal propensity to consume, d is the sensitivity of investment demand to interest rate. Affected by American financial crisis in 2007, the Chinese resident's marginal propensity to consumption and investment demand's sensitivity to interest rate both became smaller, which decides that the slope of curve IS curve increases and IS curve becomes steeper. According

to model above, IS curve in the modified M-F model has a smaller slope than that of the unmodified IS curve, so the modified IS curve becomes a bit smoother after it becomes steeper under the influence of financial crisis.

According to the above model, the slope of LM curve is k/h , and k and h are decided by the money demand function of $M_d = K_y - h r$, where K_y is money transaction demand and $-h r$ is money speculative demand. After 2008, affected by American financial crisis, economic structures were adjusted all over the world. Under such a background, Chinese growth speed has slowed down compared with those in former years. People were willing to hold any amount of currency, in which case h becomes larger and LM curve thus becomes smoother.

From the analysis above, we know that BP curve and LM curve in the modified M-F model for China during 2008 and 2012 are both smooth and IS curve is steeper than BP curve and LM curve. See Figure 1.

3.3 ANALYSIS ON THE EFFECTIVENESS OF FISCAL AND MONETARY POLICIES OF CHINESE MACRO-ECONOMY IN POST-CRISIS PERIOD

3.3.1 Analysis on Fiscal Policy

A series of fiscal policies implemented in China these years can be illustrated by Figure 13, which shows the good effects of the fiscal policy. Before the outbreak of sub-prime mortgage crisis in 2007, China was in an overheated economic growth period with foreign trade surplus. The imbalance at that time is shown in Figure 9. To control the overheated economy, China continued to implement the moderate fiscal policy. From 2008 to now, due to American sub-prime mortgage crisis, Chinese economic growth declined greatly with continuous surplus of foreign trade. The imbalance state is shown in Figure 7. To achieve economic equilibrium, IS curve moves to the right, namely implementing the positive fiscal policy, in which case there is the balance of payment deficit, the domestic currency depreciates, and the net export increases to make IS curve shift to the right continually, then BP curve also shifts to the right and reaches a new equilibrium point. The specific analysis is shown in Figure 15. It is effective to implement the positive fiscal policy in the face of such an imbalanced state. The fact was also like this. China has implemented the positive fiscal policy since 2008. Especially, after the global financial crisis caused by American sub-prime mortgage crisis, China rapidly launched a series of measures to enlarge domestic demand and implemented the positive fiscal policy, which effectively curbed the tendency of economic decline. Similarly, from 2009 to now, China has been in the imbalanced economic state shown in Figure 7 and has adopted the positive fiscal policy. These fiscal policies push the economy shifting from the former imbalanced state to a new equilibrium point to effectively increase investment, promote

industrial structure upgrading, stimulate the economic growth, change economic development mode, increase employment, and maintain the stable and rapid development of national economy.

3.3.2 Analysis on Monetary Policy

The analysis on the effect of Chinese monetary policy in recent years can be made combining Figure 11, Figure 12 and Figure 14. Because before 2002, China implemented a fixed exchange rate system under the partial-closed economic conditions and the implementation of Chinese monetary policy is relatively effective, the paper does not make a detailed analysis. Since 2003, Chinese economy has been in an incomplete open situation, but as Chinese entry into WTO, the economic openness has become bigger and bigger. According to previous analysis, China is implementing an incomplete floating exchange rate system, which is partial to the characteristics of fixed exchange rate system. Under such macroeconomic conditions, according to annual macroeconomic conditions at home and abroad, the central bank adopted a series of monetary policies according to the direction of monetary policy. The effect of monetary policy showed a decrease tendency, indicating the effect of monetary policy is characterized by disequilibrium in different economic environments. For instance, the imbalance state in 2007 due to the overheated economy and continuous surplus of trades is shown in Figure 9. According to the analyses above, to make the economy reach a new equilibrium and LM curve shift to the left, which means implementing the tight monetary policy to increase the interest rate, the securities market will meet with obstruction. In the meantime, there is international flight and the equilibrium point of international payments balance shifts to the left, in which case the international payment balance deficit leads to the depreciation of local currency. To maintain a fixed exchange rate, if keeping decreasing money supply, economic disequilibrium will be aggravated; if increasing money supply, LM will return to the position of LM_0 , maintaining a fixed exchange rate and restore economic equilibrium. However, if increasing money supply, which causes the rise in price of securities market and some international capital inflows, the trade surplus will be further enlarged. Therefore, the effect of monetary policy may be implicit or ineffective, or even uncertain. Practice has proved that the tight monetary policy was adopted in 2007, including 10 times of the increase in deposit re-serve ratio successively, the 14.5% required reserves by the end of year, 6 times of increase in interest rate, and other series of tight monetary policies, but the successive adjustment of deposit reserve ratio and the increase in interest rate had no obvious effect. The increase in interest rate can restrain the overheated investment by increasing enterprise's financing cost and the outflow cost of deposit. However, the price rising at that time was mainly promoted by the cost of agricultural products, so the

action of interest rate increase alone could not add the supply of agricultural products, but might add other production costs and become the factor to push prices higher. Therefore, the monetary policy such as interest rate rise cannot achieve the ideal effect of money deflation.

To sum up the analysis combining the modified M-F model and the reality, the inefficacy of monetary policy is mainly shown in the following aspects: the easy monetary policy fails to avoid price falling; the effect of money supply on economic growth weakens; continuous interest rate decline has little effect on stimulating consumption; the monetary policy has limited effects in promoting investment. There is limited space for the operation of monetary policy, especially under the background of global financial crisis. In addition, Chinese economy is facing internal economic structure imbalance and fierce external impact, thus having the risk of continuous decline in economic growth rate, and the central bank lacks the operation tools for an open market, so central bank's operation of monetary policy becomes more difficult.

Therefore, through the analysis in the paper with the principles of modified M-F model, the reasons for the unstable effect of Chinese monetary policy in recent years include: first, China is in a transitional period from the fixed exchange rate system to the floating exchange rate system. If we analyse the effectiveness of the monetary policy simply from the perspective of fixed exchange rate, we can see from the modified M-F model, the monetary policy may always aggravate the unstable and imbalanced situation of economy, thus increasing the difficulty in policy intervention. Analysing the effectiveness of the monetary policy simply from the perspective of floating exchange rate, the monetary policy has some effects, but with an uncertainty. Analysing the effect of the monetary policy of Chinese exchange rate system during the transitional period, we find economic intervention becomes more and more difficult, which is quite obvious in Chinese economic practice in recent years. Second, since Chinese entry into WTO, because of the trend of global integration of finance and economy, the open degree of Chinese economy has been increased constantly and the capital market such as securities and real estate began to show more and more characteristics of international capital market behaviours. On the other hand, the securities market and real estate market play more and more important roles in international capital flows and international payment balance. Therefore, implementing economic intervention in line with the market rules in the two markets should be an important part of the inter-national payment balance policy, and the effect on securities market should be taken into account when making the interest rate policy.

4 Chinese Macroeconomic Policy Orientations in the Post-crisis Period

Considering current macroeconomic background of China and basing on traditional M-F model, the paper modifies M-F model in the case that the traditional M-F model hypothesis are not suitable for Chinese macroeconomic conditions, to make the model applicable for the analysis on Chinese macroeconomic policy. The following conclusions are drawn through the analysis on the effectiveness of Chinese economic policy with the modified M-F model: under the floating exchange rate system, when the BP curve has a negative slope, the fiscal policy has significant effects and the monetary policy has effects which are uncertain to some extent; under the fixed exchange rate system, the monetary policy may aggravate the uncertainty and imbalance of economy and the increase the difficulty in policy intervention. When adopting the fiscal policy, the policy has more effects only with the positive cooperation of the monetary policy.

In conclusion, Chinese economy is experiencing the economic growth decline with a high unemployment rate and trade surplus. The economy is in an unbalanced status shown in Figure 7. Because China is in a transitional period from the fixed exchange rate to the floating ex-change rate under incomplete open economic conditions, the country should adopt the macroeconomic policy orientation dominated by the positive fiscal policy and supplemented by the monetary policy, adjust the monetary policy during implementation little by little and make further decisions considering practical effects. The adjustment process is shown in Figure 16. First to adopt the proactive fiscal policy, such as increasing government's fiscal investment in the projects under construction, the construction of indemnificatory housing, social security, employment and medical treatment, to make IS curve move to the right from IS_0 to IS_1 and the equilibrium point move from F_0 to F_1 of international payment balance deficit. Meanwhile, to keep Chinese exchange rate unchanged (based on current exchange rate system, rather than immutable), properly easy monetary policy, such as the interest reduction, the reduction of deposit reserve ratio and should be implemented for coordination to increase money supply. In this case, LM curve moves to the right from LM_0 to LM_1 and the intersection of IS_1 and LM_1 forms a new equilibrium point F_2 , very close to the intersection of BP curve and employment curve. If the adjustment is proper and the internal and external equilibriums can be reached at the same time. It can be seen that the active fiscal policy has more effects with positive coordination of monetary policy.

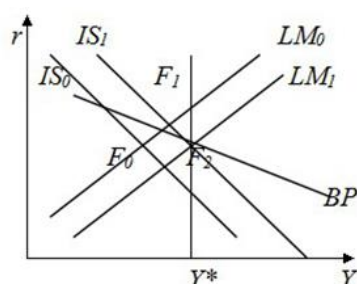


FIGURE 16 Adjustment Process from Internal and External Imbalance to Internal and External Balance

5 Conclusions

The sub-prime mortgage crisis occurred in America in 2007 caused a widespread financial crisis and profound economic crisis there and in developed economic entities. Because of its great impact on Chinese economy, China launched many fiscal and monetary policies to relieve the influence of financial crisis on economy. This paper proposed the hypotheses of M-F model applicable to current Chinese economic environment, modified the traditional M-F model and analysed the effectiveness of fiscal and monetary policies under different exchange rate systems. Conclusions are as follows: first, under the fixed exchange rate system, the monetary policy may aggravate the unstable and imbalanced situation of economy and thus increases the difficulty in policy intervention. The effects will be significant by implementing the fiscal policy with the positive coordination of monetary policy. Second, under the floating exchange rate system, the fiscal policy has significant effects, but the effects of the monetary policy are uncertain. Third, the paper uses the modified M-F to analyse the effectiveness of Chinese macro-economic policy and puts forward the policy orientation in the post-crisis period. Currently, China is in a transitional period from the fixed exchange rate to the floating exchange rate in incomplete open economic conditions and is experiencing growth rate decline, high unemployment rate and trade surplus. China should adopt the macroeconomic policy orientation dominated by the

positive fiscal policy and supplemented by the monetary policy. Furthermore, China should gradually adjust the monetary policy during implementation little by little and make further decisions considering practical effects.

Acknowledgments

In the process of writing paper, thanks for the help and guidance of my colleagues and supervisor.

References

- [1] Yiping H 2009 American financial crisis and prospect of chinese economic growth *Studies of International Finance* 12(1) 32-5 (in Chinese)
- [2] Ping L, Genqian Y 2009 Systematic review and thinking the process-impact of global financial crisis on china's economy *China Industrial Economics* 15(10) 5-22 (in Chinese)
- [3] Sanlin J 2010 On the trend of china economic development in post-crisis era *Journal of Chongqing University of Technology* 18(5) 1-3 (in Chinese)
- [4] Rui C, Lizhen L 2011 Analysis on effectiveness of china's macroeconomic policy based upon mundell-fleming model *Statistics and Decision* 102(4) 101-4 (in Chinese)
- [5] Boke J M 2005 *Macro-financial stability and safety under open condition* (1st Edition) Shanghai. Press: Fudan University Press (in Chinese)
- [6] Xiaohui L M 2008 *Choice and Transformation of RMB Exchange Rate System: From the Angle of Social Welfare* (1st Edition) Beijing. Press: Renmin Press (in Chinese)
- [7] Jun W 2006 Modification of mundell-fleming model-effect of china's economic growth on rmb exchange rate *The Journal of Quantitative & Technical Economics* 20(6) 116-23 (in Chinese)
- [8] Delei Y, Xin C 2011 Interest rate, stock index and international capital flows-theoretical model and empirical test *International Finance* 35(4) 23-61 (in Chinese)
- [9] Huaimin W 2004 An empirical analysis on china's exchange rate and export effect *Journal of Guizhou Commercial College* 68(2) 9-11 (in Chinese)
- [10] Xin C M 2009 *Foreign trade multiplier and its Influencing factors-taking china for an example Gansu* Press: Lanzhou University of Finance and Economics (in Chinese)
- [11] Research Team of Office of Foreign Economics, Institute of Macroeconomics 2007 Analysis on reasons and trends of china's trade surplus *Macroeconomic Management* 49(8) 24-6 (in Chinese)
- [12] Qingming Z 2007 Causes and treatment of growth in china's foreign exchange reserves *Studies of International Finance* 86(8) 73-9 (in Chinese)

Authors



Jianxin Bi, born in May, 1974, Jiutai City, Jilin Province, P.R. China

Current position, grades: the lecturer of Faculty of Computer and Information, Zhejiang Wanli University.

University studies: doctoral candidate of Faculty of Management, the University of Shanghai for Science and Technology.

Scientific interest: Financial and fiscal Management, data mining.

Publications: Presided over 3 scientific research projects the completion of provincial; more than 10 papers published in various journals.

Experience: Graduated from Fu zhou University in 2005, has completed 3 scientific research projects; more than 10 papers published in various journals.



Lianghai Lei, born in February, 1962, Qianjiang City, Hubei Province, P.R. China

Current position, grades: professor, doctoral supervisor and Doctor of Economics of Faculty of Management, the University of Shanghai for Science and Technology.

University studies: Graduated from Shanghai University of Finance and Economics in 1996, received a doctor's degree in Economics.

Scientific interest: Financial and fiscal Management, Company financial, The theory and methods of enterprise internal control.

Publications: Presided over the completion of national and provincial, than 20 scientific research projects; more than 140 papers published in various journals.

Experience: Graduated from Shanghai University of Finance and Economics in 1996, received a doctor's degree in economics, was approved as a tutor of doctoral students in 2002; has completed including the national and provincial and ministerial level, than 20 scientific research projects; more than 140 papers published in various journals published monographs, 7 teaching materials.

Approximation reductions in an incomplete variable precision multigranulation rough set

Sheng Yao, Longshu Li^{1*}

¹ School of Computer Science and Technology, Anhui University, 111 Jiu Long Road, Hefei, China

Received 1 March 2014, www.tsi.lv

Abstract

This paper deals with approaches to the granular space reductions in the variable precision multigranulation rough set model. The main objective of this study is to extend four kinds of the granular space reductions called a tolerance relations optimistic multigranulation β lower approximation distribution reduction, a tolerance relations optimistic multigranulation β upper approximation distribution reduction, a tolerance relations pessimistic multigranulation β lower approximation distribution reduction and a tolerance relations pessimistic multigranulation β upper approximation distribution reduction, which preserve the optimistic/pessimistic multigranulation β lower/upper approximation distribution of the decision classes. Some judgement theorems are investigated. The example proves that the new variable precision multigranulation rough set model can effectively deal with incomplete information, from which we can obtain approaches to the granular space reductions of incomplete decision systems in variable precision multigranulation rough theory.

Keywords: Approximation Reduction, Tolerance Relation, Variable Precision Rough Set, Multigranulation

1 Introduction

In the past twenty years, rough set theory [1], which was first proposed by Pawlak, has been widely applied in the field of data mining [2-5], feature selection [20], machine learning, decision support [6-8], pattern recognition [9, 10] and so on. One of the important aspects of this theory is searching for particular subsets of condition granulations, which have the same information for classification purposes as all the condition granulations.

From the viewpoint of the granular computing, an equivalence relation can be regarded as a granulation. So the classical rough set can be regarded as a single granulation rough set because it is constructed on only one equivalence relation. However, in fact, we often describe an obscure concept by multiple granulations not by one single granulation for satisfying the demand of users or problem specification. Based on this need, Qian et al. proposed multigranulation rough set theory [11], in which an obscure concept is described by multiple equivalence relations on one universe. Later, some extension to multigranulation rough set has been appeared rapidly. Qian [12] proposed a multigranulation rough set model based on tolerance relation, in which an incomplete decision table was researched under multiple tolerance relations. The topological properties of multigranulation rough sets was first researched by Raghavan and Tripathy [13]. Xu et al. [14-16] developed a variable multigranulation rough set model, a fuzzy

multigranulation rough set model and an ordered multigranulation rough set model. Variable Precision multigranulation rough set was proposed and has been studied by many researchers.

In recent years, from the classical rough set model to the multigranulation rough set model, more attention has been paid to the granular space reduction. In a single granulation rough set model, many types of attribute reduction have been studied. β -reduct was researched in the variable precision rough set proposed by Ziarko [17]. The heuristic method has been applied in the attribute reduction of all the kinds of single rough set model according to different requirements. In addition, discernibility matrix is also used to eliminate the redundant condition attributes. S Kowron [18] first proposed famous knowledge reduction algorithm based on discernibility matrix. Later, many researchers have different improvements on discernibility matrix for getting better attribute reduction [21].

At present, in many kinds of multigranulation rough set model, the heuristic approach is usually used to the granular space reduction. However, the discernibility matrix is still not be researched for the granular space reduction in the context of multigranulation, which limits the development of the granular space reduction in the multigranulation rough set.

The objective of this paper is to introduce four concepts of approximation reduction named as a tolerance relations optimistic multigranulation β lower

* Corresponding author - Tel: +86-0551-63866273; fax: +86-0551-63866273; E-mail: fisheryao@126.com

approximation distribution reduction, a tolerance relations optimistic multigranulation β upper approximation distribution reduction, a tolerance relations pessimistic multigranulation β lower approximation distribution reduction and a tolerance relations pessimistic multigranulation β upper approximation distribution reduction, which preserve the optimistic/pessimistic multigranulation β lower/upper approximation distribution of the decision classes. The rest is organized as follows. Some preliminary concepts such as rough set, incomplete decision system, variable precision rough set and multigranulation rough sets are briefly reviewed in Section 2. In Section 3, the notions of a tolerance relations optimistic multigranulation β lower approximation distribution reduction, a tolerance relations optimistic multigranulation β upper approximation distribution reduction, a tolerance relations pessimistic multigranulation β lower approximation distribution reduction and a tolerance relations pessimistic multigranulation β upper approximation distribution reduction are introduced to an incomplete decision system in the content of multigranulation, and their some important properties are also obtained. In Section 4, the approaches to the four approximation reductions are provided. An illustrate example is employed to examine their validity in Section 5. In Section 6, we also compare the performances between the granular space reduction based on heuristic approach and the proposed discernibility matrix approach through experiment. Finally, Section 7 concludes this paper by giving some discussions.

2 Preliminaries

In this section, we briefly review some basic concepts such as rough set, incomplete decision system, variable precision rough set and multigranulation rough sets.

2.1 PAWLAK ROUGHT SET

Definition 1. [1] Let $K=(U,R)$ be an approximation space, where U is a finite and nonempty set of objects called the universe, and $R \subseteq U \times U$ is an equivalence relation on U . U/R denote the partition of U induced by R , which is the set of equivalence classes generated by R . For any set $X \subseteq U$, the lower approximation $\underline{R}(X)$ and the upper approximation $\overline{R}(X)$ are defined by [12]:

$$\underline{R}(X) = \bigcup \{ [x] \in U/R \mid [x]_R \subseteq X \}, \tag{1}$$

$$\overline{R}(X) = \bigcup \{ [x] \in U/R \mid [x]_R \cap X \neq \emptyset \}. \tag{2}$$

2.2 TOLERANCE RELATION

Definition 2. [19] Let $I=(U,AT,V,f)$ be an incomplete information system, $B \subseteq AT$ an attribute set. We denote some null value by *, a tolerance relation on U is defined as follows:

$$TOL(B) = \left\{ \begin{aligned} &(x,y) \in U \times U : \forall a \in B, a(x) \\ &= a(y) \vee a(x) = * \vee a(y) = * \end{aligned} \right\}. \tag{3}$$

2.3 VARIABLE PRECISION ROUGH SET

Definition 3. [17] Let U be a finite nonempty universal set, for two arbitrary nonempty sets X and Y , the relative degree of misclassification of the set X with respect to the set Y is defined as follows:

$$c(X,Y) = \begin{cases} 1 - \text{card}(X \cap Y) / \text{card}(X), & \text{card}(X) > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

where card denotes set cardinality.

2.4 MULTIGRANULATION ROUGH SET

2.4.1 Optimistic multigranulation rough set

Definiton 4. [11] Let $I=(U,AT,V,f)$ be an information system, in which $A_1, A_2, \dots, A_m \subseteq AT$, and $X \subseteq U$. The optimistic multigranulation lower and upper approximations are denoted by $\sum_{i=1}^m \overset{o}{A_i}(X)$ and

$$\overline{\sum_{i=1}^m \overset{o}{A_i}(X)}, \text{ respectively,}$$

$$\sum_{i=1}^m \overset{o}{A_i}(X) = \left\{ x \in U : [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \dots \vee [x]_{A_m} \subseteq X \right\}, \tag{5}$$

$$\overline{\sum_{i=1}^m \overset{o}{A_i}(X)} = \sim \left(\overline{\sum_{i=1}^m \overset{o}{A_i}(\sim X)} \right), \tag{6}$$

where $[x]_{A_i}$ ($i=1,2,\dots,m$) is the equivalence class of x in terms of set of attributes A_i , and $\sim X$ is the complement of X .

2.4.2 Pessimistic multigranulation rough set

Definiton 5. [11] Let $I=(U,AT,V,f)$ be an information system, in which $A_1, A_2, \dots, A_m \subseteq AT$, and $X \subseteq U$. The pessimistic multigranulation lower and upper

approximations are denoted by $\sum_{i=1}^m A_i(X)$ and

$$\overline{\sum_{i=1}^m A_i(X)}, \text{ respectively,}$$

$$\sum_{i=1}^m A_i(X) = \{x \in U : [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \dots \vee [x]_{A_m} \subseteq X\}, \quad (7)$$

$$\overline{\sum_{i=1}^m A_i(X)} = \sim \left(\sum_{i=1}^m A_i(\sim X) \right), \quad (8)$$

where $[x]_{A_i}$ ($i=1,2,\dots,m$) is the equivalence class of x in terms of set of attributes A_i , and $\sim X$ is the complement of X .

3 Approximation Distribution Reduction

Definition 6. Suppose $S = \langle U, C \cup D, f \rangle$ is an incomplete decision information system, where $C = \{c_1, c_2, \dots, c_n\}$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, then for any $B = \{b_1, b_2, \dots, b_m\} \subseteq C$, the tolerance relations optimistic multigranulation β lower approximation distribution function with respect to B , the tolerance relations optimistic multigranulation β upper approximation distribution function with respect to B , the tolerance relations pessimistic multigranulation β lower approximation distribution function with respect to B and the tolerance relations pessimistic multigranulation β upper approximation distribution function with respect to B are denoted by $LB_\beta^{O,T}(D)$, $UB_\beta^{O,T}(D)$, $LB_\beta^{P,T}(D)$ and $UB_\beta^{P,T}(D)$, respectively, where

$$LB_\beta^{O,T}(D) = \left\{ \sum_{i=1}^m b_i^{O,T}(D_1), \sum_{i=1}^m b_i^{O,T}(D_2), \dots, \sum_{i=1}^m b_i^{O,T}(D_r) \right\}, \quad (9)$$

$$UB_\beta^{O,T}(D) = \left\{ \overline{\sum_{i=1}^m b_i^{O,T}(D_1)}, \overline{\sum_{i=1}^m b_i^{O,T}(D_2)}, \dots, \overline{\sum_{i=1}^m b_i^{O,T}(D_r)} \right\}, \quad (10)$$

$$LB_\beta^{P,T}(D) = \left\{ \sum_{i=1}^m b_i^{P,T}(D_1), \sum_{i=1}^m b_i^{P,T}(D_2), \dots, \sum_{i=1}^m b_i^{P,T}(D_r) \right\}, \quad (11)$$

$$UB_\beta^{P,T}(D) = \left\{ \overline{\sum_{i=1}^m b_i^{P,T}(D_1)}, \overline{\sum_{i=1}^m b_i^{P,T}(D_2)}, \dots, \overline{\sum_{i=1}^m b_i^{P,T}(D_r)} \right\}. \quad (12)$$

By the four approximation distribution functions, we have the following definition of the four corresponding approximation consistent sets and the four corresponding approximation reduct in incomplete decision information system.

Definition 2. Suppose $S = \langle U, C \cup D, f \rangle$ is an incomplete decision information system, where $C = \{c_1, c_2, \dots, c_n\}$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, then for any $B = \{b_1, b_2, \dots, b_m\} \subseteq C$.

(1) If $LB_\beta^{O,T}(D) = LC_\beta^{O,T}(D)$, we say that B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S . If B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set, and no proper subset of B is tolerance relations optimistic multigranulation β lower approximation distribution consistent, then B is called a tolerance relations optimistic multigranulation lower approximation distribution reduct of C .

(2) If $UB_\beta^{O,T}(D) = UC_\beta^{O,T}(D)$, we say that B is a tolerance relations optimistic multigranulation β upper approximation distribution consistent attributes set of S . If B is a tolerance relations optimistic multigranulation β upper approximation distribution consistent attribute set, and no proper subset of B is tolerance relations optimistic multigranulation β upper approximation distribution consistent, then B is called a tolerance relations optimistic multigranulation upper approximation distribution reduct of C .

(3) If $LB_\beta^{P,T}(D) = LC_\beta^{P,T}(D)$, we say that B is a tolerance relations pessimistic multigranulation β lower approximation distribution consistent attributes set of S . If B is a tolerance relations pessimistic multigranulation β lower approximation distribution consistent attribute set, and no proper subset of B is tolerance relations pessimistic multigranulation β lower approximation distribution consistent, then B is called a tolerance relations pessimistic multigranulation lower approximation distribution reduct of C .

(4) If $UB_\beta^{P,T}(D) = UC_\beta^{P,T}(D)$, we say that B is a tolerance relations pessimistic multigranulation β upper approximation distribution consistent attributes set of S . If B is a tolerance relations pessimistic multigranulation β upper approximation distribution consistent attribute set, and no proper subset of B is tolerance relations pessimistic multigranulation β upper approximation distribution consistent, then B is called a tolerance relations pessimistic multigranulation upper approximation distribution reduct of C .

From Definition 2, we obtain the following interpretations:

- (1) A tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S is a subset of the attributes, which preserves the optimistic multigranulation β lower approximations of each class; a tolerance relations optimistic multigranulation β lower approximation distribution reduct of S is a minimal subset of the attributes, which preserves the optimistic multigranulation β lower approximations of each decision class.
- (2) A tolerance relations optimistic multigranulation β upper approximation distribution consistent attributes set of S is a subset of the attributes, which preserves the optimistic multigranulation β upper approximations of each class; a tolerance relations optimistic multigranulation β upper approximation distribution reduct of S is a minimal subset of the attributes, which preserves the optimistic multigranulation β upper approximations of each decision class.
- (3) A tolerance relations pessimistic multigranulation β lower approximation distribution consistent attributes set of S is a subset of the attributes, which preserves the pessimistic multigranulation β lower approximations of each class; a tolerance relations pessimistic multigranulation β lower approximation distribution reduct of S is a minimal subset of the attributes, which preserves the pessimistic multigranulation β lower approximations of each decision class.
- (4) A tolerance relations pessimistic multigranulation β upper approximation distribution consistent attributes set of S is a subset of the attributes, which preserves the pessimistic multigranulation β upper approximations of each class; a tolerance relations pessimistic multigranulation β upper approximation distribution reduct of S is a minimal subset of the attributes, which preserves the pessimistic multigranulation β upper approximations of each decision class.

For convenience of discussions, we first give some equivalent characterizations for four kinds of approximation distribution consistent attributes sets.

Let $S = \langle U, C \cup D, f \rangle$ be an incomplete decision information system, where $C = \{c_1, c_2, \dots, c_n\}$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$ then for any $B = \{b_1, b_2, \dots, b_m\} \subseteq C$, denote

$$OL_{\beta}^{c,T}(x) = \left\{ D_j \in U/IND(\{d\}) : x \in \sum_{i=1}^n \frac{c_i}{\beta} (D_j) \right\}, \quad (13)$$

$$OU_{\beta}^{c,T}(x) = \left\{ D_j \in U/IND(\{d\}) : x \in \sum_{i=1}^n \frac{c_i}{\beta} (D_j) \right\}, \quad (14)$$

$$PL_{\beta}^{c,T}(x) = \left\{ D_j \in U/IND(\{d\}) : x \in \sum_{i=1}^n \frac{c_i}{\beta} (D_j) \right\}, \quad (15)$$

$$PU_{\beta}^{c,T}(x) = \left\{ D_j \in U/IND(\{d\}) : x \in \sum_{i=1}^n \frac{c_i}{\beta} (D_j) \right\}. \quad (16)$$

Theorem 1. Let $S = \langle U, C \cup D, f \rangle$ be an incomplete decision information system, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, $\forall B \subseteq C$, the following properties hold.

$$\sum_{i=1}^n \frac{c_i}{\beta} (x) = \left\{ D_j : x \in U, D_j \in OL_{\beta}^{c,T}(x), j=1, 2, \dots, r \right\}, \quad (17)$$

$$\sum_{i=1}^n \frac{c_i}{\beta} (x) = \left\{ D_j : x \in U, D_j \in OU_{\beta}^{c,T}(x), j=1, 2, \dots, r \right\}, \quad (18)$$

$$\sum_{i=1}^n \frac{c_i}{\beta} (x) = \left\{ D_j : x \in U, D_j \in PL_{\beta}^{c,T}(x), j=1, 2, \dots, r \right\}, \quad (19)$$

$$\sum_{i=1}^n \frac{c_i}{\beta} (x) = \left\{ D_j : x \in U, D_j \in PU_{\beta}^{c,T}(x), j=1, 2, \dots, r \right\}. \quad (20)$$

Proof. From the above equivalent characterizations, these are straightforward.

In what follows, we study some judgement methods of the four kinds of approximation consistent set in an incomplete decision.

Theorem 2 (Judgement theorem of consistent set I) Suppose $S = \langle U, C \cup D, f \rangle$ is an incomplete decision information system, where $C = \{c_1, c_2, \dots, c_n\}$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, then for any $B = \{b_1, b_2, \dots, b_m\} \subseteq C$, we have:

- (1) B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S iff $\forall x \in U, OL_{\beta}^{B,T}(x) = OL_{\beta}^{c,T}(x)$;

(2) B is a tolerance relations optimistic multigranulation β upper approximation distribution consistent attributes

set of S iff $\forall x \in U, OU_{\beta}^{B,T}(x) = OU_{\beta}^{C,T}(x)$;

(3) B is a tolerance relations pessimistic multigranulation β lower approximation distribution consistent attributes set of S iff

$\forall x \in U, PL_{\beta}^{B,T}(x) = PL_{\beta}^{C,T}(x)$;

(4) B is a tolerance relations pessimistic multigranulation β upper approximation distribution consistent attributes set of S iff

$\forall x \in U, PU_{\beta}^{B,T}(x) = PU_{\beta}^{C,T}(x)$.

Proof. We only prove (1), others can be proved analogously.

(1) Sufficiency: The assumption that B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S implies $LB_{\beta}^{O,T}(D) = LC_{\beta}^{O,T}(D)$. Then by equation (7), we have

$$\sum_{i=1}^m b_i(D_j) = \sum_{i=1}^n c_i(D_j) \text{ for each } D_j \in IND(\{d\}), \text{ it}$$

follows that $\forall x \in U, x \in \sum_{i=1}^m b_i(D_j) \Leftrightarrow x \in \sum_{i=1}^n c_i(D_j)$ and

$$x \notin \sum_{i=1}^m b_i(D_j) \Leftrightarrow x \notin \sum_{i=1}^n c_i(D_j), \text{ i.e.}$$

$$\forall x \in U, OL_{\beta}^{B,T}(x) = OL_{\beta}^{C,T}(x).$$

Necessary: Since $\forall x \in U, OL_{\beta}^{B,T}(x) = OL_{\beta}^{C,T}(x)$, then

$$\forall x \in U, x \in \sum_{i=1}^m b_i(D_j) \Leftrightarrow x \in \sum_{i=1}^n c_i(D_j) \text{ and}$$

$$x \notin \sum_{i=1}^m b_i(D_j) \Leftrightarrow x \notin \sum_{i=1}^n c_i(D_j), \text{ it follows that}$$

$$\sum_{i=1}^m b_i(D_j) = \sum_{i=1}^n c_i(D_j), \text{ for each } D_j \in IND(\{d\}),$$

i.e. $LB_{\beta}^{O,T}(D) = LC_{\beta}^{O,T}(D)$. Consequently, B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S .

Theorem 2 provides an approach to judge whether a subset of condition attributes is a tolerance relations optimistic/pessimistic multigranulation β lower/upper approximation distribution consistent attributes set of S .

4 Approaches to approximation Distribution Reduction

In this section, we further provide practical approaches to approximation distribution reductions in an incomplete decision information system.

First, we give the following notions:

Definition 7. Let $S = \langle U, C \cup D, f \rangle$ be an incomplete

decision information system, where $C = \{c_1, c_2, \dots, c_n\}$,

$0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, then for each

$B = \{b_1, b_2, \dots, b_m\} \subseteq C$, we denote

$$D_{OL}^{*\beta} = \{(x, y) : D_j \in OL_{\beta}^{C,T}(x), y \notin D_j, (x, y) \in U \times U \wedge (x, y) \notin T^{\beta}(C)\}, \quad (21)$$

$$D_{OU}^{*\beta} = \{(x, y) : x \in D_j, D_j \notin OU_{\beta}^{C,T}(y), (x, y) \in U \times U \wedge (x, y) \notin T^{\beta}(C)\}, \quad (22)$$

$$D_{PL}^{*\beta} = \{(x, y) : D_j \in PL_{\beta}^{C,T}(x), y \notin D_j, (x, y) \in U \times U \wedge (x, y) \notin T^{\beta}(C)\}, \quad (23)$$

$$D_{PU}^{*\beta} = \{(x, y) : x \in D_j, D_j \notin PU_{\beta}^{C,T}(y), (x, y) \in U \times U \wedge (x, y) \notin T^{\beta}(C)\}. \quad (24)$$

Defined by

$$D_l^{\beta}(x, y) = \begin{cases} a \in C : f_a(x) \neq f_a(y) \wedge f_a(x) \neq * \wedge f_a(y) \neq *, & (x, y) \in D_l^{*\beta} \\ \phi, & (x, y) \notin D_l^{*\beta} \end{cases}$$

($l \in \{OL, OU, PL, PU\}$), then

$D_l^{\beta}(x, y), l \in \{OL, OU, PL, PU\}$ are called the tolerance relations optimistic multigranulation β lower approximation distribution discernibility attribute set, the tolerance relations optimistic multigranulation β upper approximation distribution discernibility attribute set, the tolerance relations pessimistic multigranulation β lower approximation distribution discernibility attribute set and the tolerance relations pessimistic multigranulation β upper approximation distribution discernibility attribute set, respectively.

According to Definition 7, the following judgement theorem of a consistent set can be obtained.

Theorem 3 (Judgement theorem of consistent set II)

Suppose $S = \langle U, C \cup D, f \rangle$ is an incomplete decision

information system, where $C = \{c_1, c_2, \dots, c_n\}$,

$B = \{b_1, b_2, \dots, b_m\} \subseteq C$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$, then

(1) B is a tolerance relations optimistic multigranulation β lower approximation distribution consistent attributes set of S iff $B \cap D_{OL}^{\beta}(x, y) \neq \phi$ for each $(x, y) \in D_{OL}^{*\beta}$;

(2) B is a tolerance relations optimistic multigranulation β upper approximation distribution consistent attributes set of S iff $B \cap D_{OU}^{\beta}(x, y) \neq \phi$ for each $(x, y) \in D_{OU}^{*\beta}$;

(3) B is a tolerance relations pessimistic multigranulation β lower approximation distribution consistent attributes set of S iff $B \cap D_{PL}^{\beta}(x, y) \neq \phi$ for each $(x, y) \in D_{PL}^{*\beta}$;

(4) B is a tolerance relations pessimistic multigranulation β upper approximation distribution

consistent attributes set of S iff $B \cap D_{PU}^\beta(x, y) \neq \emptyset$ for each $(x, y) \in D_{PU}^{*\beta}$.

Proof. We only prove (1), others can be proved analogously.

(2) Sufficiency: Suppose $\exists(x, y) \in D_{OL}^{*\beta}$ such that $B \cap D_{OL}^\beta(x, y) = \emptyset$, and then we have $(x, y) \in T^\beta(B)$, i.e. $y \in T_B^\beta(x)$. Since $(x, y) \in D_{OL}^{*\beta}$, then there must be $D_j \in U/D$ such that $x \in \bigcup_{i=1}^n c_i^{o,T} (D_j)$ and $y \notin D_j$.

By condition we know that B is a tolerance relations optimistic multigranulation β lower approximate distribution consistent attributes set of S , then $x \in \bigcup_{i=1}^m b_i^{o,T} (D_j)$, i.e. there exist $b_k \in B$ such that $T_{\{b_k\}}^\beta(x) \subseteq D_j$. By the basic property of the tolerance relation, we have $T_B^\beta(x) \subseteq T_{\{b_k\}}^\beta(x) \subseteq D_j$, from which we can conclude that $y \in D_j$, which is contradictive to the assumption $(x, y) \in D_{OL}^{*\beta}$ because $(x, y) \in D_{OL}^{*\beta} \Rightarrow y \notin D_j$.

Necessary: Since $B \subseteq C$, then we have $OL_\beta^{B,T}(x) \subseteq OL_\beta^{C,T}(x)$ for each $x \in U$ obviously. Therefore, it must be proved that $OL_\beta^{C,T}(x) \subseteq OL_\beta^{B,T}(x)$ for each $x \in U$.

If $\exists(x, y) \in D_{OL}^{*\beta} (D_j \in OL_\beta^{C,T}(x), y \notin D_j)$ such that $B \cap D_{OL}^\beta(x, y) = \emptyset$, then we have $(x, y) \in T^\beta(B)$, i.e. $y \in T_B^\beta(x)$. By the basic property of the tolerance relation, we have $y \in T_{\{b_k\}}^\beta(x)$ for each $b_k \in B$, i.e. $T_{\{b_k\}}^\beta(x) \not\subseteq D_j$ for each $b_k \in B$, it follows that $x \notin \bigcup_{i=1}^n c_i^{o,T} (D_j)$, $D_j \in OL_\beta^{B,T}(x)$. From discussions above, we can conclude that if $B \cap D_{OL}^\beta(x, y) \neq \emptyset$ for each $(x, y) \in D_{OL}^{*\beta}$, then we have $x \in \bigcup_{i=1}^n c_i^{o,T} (D_j) (D_j \in OL_\beta^{B,T}(x))$, where $D_j \in OL_\beta^{C,T}(x)$, i.e. $OL_\beta^{C,T}(x) \subseteq OL_\beta^{B,T}(x)$.

Definition 4. Let $S = \langle U, C \cup D, f \rangle$ be an incomplete decision information system, in which $C = \{c_1, c_2, \dots, c_n\}$, $B = \{b_1, b_2, \dots, b_m\} \subseteq C$, $0 \leq \beta < 0.5$, $D_l^\beta = \{D_l^\beta(x, y) : x, y \in U\}$, $l = \{OL, OU, PL, PU\}$ are called the tolerance relations optimistic multigranulation β lower approximation

distribution discernibility matrix of S , the tolerance relations optimistic multigranulation β upper approximation distribution discernibility matrix of S , the tolerance relations pessimistic multigranulation β lower approximation distribution discernibility matrix of S , and the tolerance relations pessimistic multigranulation β upper approximation distribution discernibility matrix of S , respectively. Let

$$\Delta_l^\beta = \bigwedge_{D_l^\beta(x,y) \in D_l^\beta} (\vee D_l^\beta(x, y)), l \in \{OL, OU, PL, PU\}.$$

Then, $\Delta_l^\beta, l \in \{OL, OU, PL, PU\}$ are, respectively, referred to as the tolerance relations optimistic multigranulation β lower approximation distribution discernibility function, the tolerance relations optimistic multigranulation β upper approximation distribution discernibility function, the tolerance relations pessimistic multigranulation β lower approximation distribution discernibility function, and the tolerance relations pessimistic multigranulation β upper approximation distribution discernibility function.

Using Boolean Reasoning technique, it is easy to obtain the following theorem by Theorem 3.

Theorem 4 (Judgement theorem of consistent set III) Suppose $S = \langle U, C \cup D, f \rangle$ is an incomplete decision information system, where $C = \{c_1, c_2, \dots, c_n\}$, $B = \{b_1, b_2, \dots, b_m\} \subseteq C$, $0 \leq \beta < 0.5$, $U/D = \{D_1, D_2, \dots, D_r\}$. The minimal disjunctive normal form of each discernibility function $\Delta_l^\beta (l \in \{OL, OU, PL, PU\})$ is

$$\Delta_l^\beta = \bigvee_{k=1}^t \left(\bigwedge_{s=1}^{q_k} a_{i_s} \right), l \in \{OL, OU, PL, PU\}, \text{ let } B_k = \{a_{i_s} : s=1, 2, \dots, q_k\}$$

and then $\{B_k : k = 1, 2, \dots, t\} (l \in \{OL, OU, PL, PU\})$ are, respectively, the set of the tolerance relations optimistic multigranulation β lower approximation distribution reduction, the set of the tolerance relations optimistic multigranulation β upper approximation distribution reduction, the set of the tolerance relations pessimistic multigranulation β lower approximation distribution reduction, the set of the tolerance relations pessimistic multigranulation β upper approximation distribution reduction.

Proof. For any $k \leq t$ and $(x, y) \in D_l^{*\beta}$, by the definition of the minimal disjunctive normal form, we have $B_k \cap D_l^\beta(x, y) \neq \emptyset$, and then from Theorem 3, we obtain the conclusion that B_k is the corresponding approximation distribution consistent attributes set.

Simultaneously, $\Delta_l^\beta = \bigvee_{k=1}^t B_k$, if B_k is formed by deleting an element from B_k , then there exist

$(x, y) \in D_i^{\beta}$ such that $B_{ik} \cap D_i^{\beta}(x, y) = \emptyset$, so B_{ik} is not an corresponding approximation distribution consistent attributes set. Hence, B_{ik} is the corresponding approximation distribution reduction.

Since all $D_i^{\beta}(x, y)$ are included in the corresponding approximation distribution discernibility function, so there does not exist other corresponding approximation distribution reduction.

Theorem 4 provides practical approaches to some attribute reductions in incomplete information decision system.

5 Example Analysis

In this section, an illustrate example is employed to explain the approach introduced in Section 3. Suppose the following incomplete information decision table is given in Table 1.

TABLE 1 An incomplete information decision table

U	a ₁	a ₂	a ₃	a ₄	d
x ₁	3	2	1	0	Φ
x ₂	2	3	2	0	Φ
x ₃	2	3	2	0	Ψ
x ₄	*	2	*	1	Φ
x ₅	*	2	*	1	Ψ
x ₆	2	3	2	1	Ψ
x ₇	3	*	*	3	Φ
x ₈	*	0	0	*	Ψ
x ₉	3	2	1	3	Ψ
x ₁₀	1	*	*	*	Φ
x ₁₁	*	2	*	*	Ψ
x ₁₂	3	2	1	*	Φ

The decision classes of objects are $D_{\Phi} = \{x_1, x_2, x_4, x_7, x_{10}, x_{12}\}$ and $D_{\Psi} = \{x_3, x_5, x_6, x_8, x_9, x_{11}\}$.

It can easily be calculated that $T_{R_c}(x_1) = \{x_1, x_{11}, x_{12}\}$,

$T_{R_c}(x_2) = \{x_2, x_3\}$, $T_{R_c}(x_3) = \{x_2, x_3\}$.

$T_{R_c}(x_4) = \{x_4, x_5, x_{10}, x_{11}, x_{12}\}$,

$T_{R_c}(x_5) = \{x_4, x_5, x_{10}, x_{11}, x_{12}\}$,

$T_{R_c}(x_6) = \{x_6\}$,

$T_{R_c}(x_7) = \{x_7, x_8, x_9, x_{11}, x_{12}\}$,

$T_{R_c}(x_8) = \{x_7, x_8, x_{10}\}$,

$T_{R_c}(x_9) = \{x_7, x_9, x_{11}, x_{12}\}$,

$T_{R_c}(x_{10}) = \{x_4, x_5, x_8, x_{10}, x_{11}\}$,

$T_{R_c}(x_{11}) = \{x_1, x_4, x_5, x_7, x_9, x_{10}, x_{11}, x_{12}\}$,

$T_{R_c}(x_{12}) = \{x_1, x_4, x_5, x_7, x_9, x_{11}, x_{12}\}$.

Thus, for $\beta = 0.3$,

$$D_{OL}^{*0.3} = \{(x_2, x_1), (x_2, x_4), (x_2, x_7), (x_2, x_{10}), (x_2, x_{12}), (x_3, x_1), (x_3, x_4), (x_3, x_7), (x_3, x_{10}), (x_3, x_{12}), (x_6, x_1), (x_6, x_2), (x_6, x_4), (x_6, x_7), (x_6, x_{10}), (x_6, x_{12})\}$$

Hence, we obtain the tolerance relations optimistic multigranulation 0.3 lower approximation distribution discernibility matrix as follows:

φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
a ₁ a ₂ a ₃	φ	φ	a ₂ a ₄	φ	φ	a ₁ a ₄	φ	φ	a ₁	φ	a ₁ a ₂ a ₃
a ₁ a ₂ a ₃	φ	φ	a ₂ a ₄	φ	φ	a ₁ a ₄	φ	φ	a ₁	φ	a ₁ a ₂ a ₃
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
a ₁ a ₂ a ₃ a ₄	a ₄	φ	a ₂	φ	φ	a ₁ a ₄	φ	φ	a ₁	φ	a ₁ a ₂ a ₃
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ

Then the tolerance relations optimistic multigranulation 0.3 lower approximation distribution discernibility function is

$$\Delta_{OL}^{0.3} = (a_1 \vee a_2 \vee a_3) \vee (a_2 \vee a_4) \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) \vee (a_1 \vee a_2 \vee a_3) \vee (a_2 \vee a_4) \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) \vee (a_1 \vee a_2 \vee a_3 \vee a_4) \vee a_4 \vee a_2 \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) = a_1 \wedge a_2 \wedge a_4.$$

In addition, we also obtain the tolerance relations optimistic multigranulation 0.3 upper approximation distribution discernibility matrix as follows:

φ	a ₁ a ₂ a ₃	a ₁ a ₂ a ₃	φ	φ	a ₁ a ₂ a ₃ a ₄	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	a ₄	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	a ₂ a ₄	a ₂ a ₄	φ	φ	a ₂	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	a ₁ a ₄	a ₁ a ₄	φ	φ	a ₁ a ₄	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	a ₁	φ	φ	a ₁	φ	φ	φ	φ	φ	φ	φ
φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
φ	a ₁ a ₂ a ₃	a ₁ a ₂ a ₃	φ	φ	a ₁ a ₂ a ₃	φ	φ	φ	φ	φ	φ

Then the tolerance relations optimistic multigranulation 0.3 lower approximation distribution discernibility function is

$$\Delta_{OL}^{0.3} = (a_1 \vee a_2 \vee a_3) \vee (a_2 \vee a_4) \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) \vee (a_1 \vee a_2 \vee a_3) \vee (a_2 \vee a_4) \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) \vee (a_1 \vee a_2 \vee a_3 \vee a_4) \vee a_4 \vee a_2 \vee (a_1 \vee a_4) \vee a_1 \vee (a_1 \vee a_2 \vee a_3) = a_1 \wedge a_2 \wedge a_4.$$

By theorem 4, $\{a_1, a_2, a_4\}$ is the tolerance relations optimistic multigranulation 0.3 lower approximation distribution reduction.

Similarly, it is not difficult to obtain that the tolerance relations pessimistic multigranulation 0.3 lower approximation distribution reduction and the tolerance relations pessimistic multigranulation 0.3 upper approximation distribution reduction in Table 1 is the set of the attributes $C = \{a_1, a_2, a_3, a_4\}$.

6 Experimental Comparison

In what follows, through experimental analysis, in the multigranulation environment, we will show the power of our proposed method by comparing it with the granular space reduction based on Heuristic approach proposed in literature [12]. Six public data sets have been downloaded from UCI Repository of Machine Learning databases. The property of these seven data sets is shown in Table 2.

TABLE 2 Data Sets Description

ID	Data Sets	Samples	Features	Decision Classes
1	Monk's Problems	432	7	2
2	Abcsic Acid Signaling Network	300	43	2
3	University	285	17	2
4	Pittburgh Bridges	108	13	2
5	Mammographic Mass	961	6	2
6	Led17	2000	22	2
7	SkillCraft1 Master Table	3395	20	2

TABLE 3 The tolerance relations optimistic multigranulation 0.2 lower approximation distribution reductions

Data ID	Reduction numbers	Granules of the shortest reduction	Granules of the longest reduction	The shortest reduction efficiency	The longest reduction efficiency
1	1	5	5	28.6%	28.6%
2	373	12	17	72.1%	60.5%
3	285	11	13	35.3%	23.5%
4	38	6	8	53.8%	38.5%
5	6	6	6	0	0
6	117	18	18	18.2%	18.2%
7	536	13	14	53.8%	0.3%

TABLE 4 Time comparisons between the two algorithms

Data ID	Time consuming of AGSR/(s)	Time consuming of MAGSR/(s)
1	45.1025	2.7641
2	37.7021	3.2001
3	24.8925	1.7541
4	0.2572	0.0852
5	384.1362	4.2531
6	1015.5264	1009.4907
7	3426.1726	3426.8823

Table 3 displays the tolerance relations optimistic multigranulation 0.2 lower approximation distribution reductions on seven public data sets. The time consuming of the two algorithms are shown in Table 4. In table 4, the granular space reduction based on Heuristic approach is denoted by HAGSR, and the proposed discernibility matrix approach is denoted by MAGSR.

It can be seen from Table 3 that the Monk's Problem data set has one reduction only, the Abcsic Acid Signaling Network data set has 373 reductions, the University data set has 285 reductions, the data set has 38 reductions, the data set has 6 reductions, the Led17 data set has 117 reductions, and the SkillCraft1 Master Table data set has These results show that the proposed discernibility matrix approach can obtain all the tolerance relations optimistic/pessisitic multigranulation 0.2 lower/upper approximation distribution reductions, but not a single corresponding reduction.

From Table 4, we can see that the time consuming of the proposed discernibility matrix approach is far lower than the time consuming of the granular space reduction based on Heuristic approach, especially to small-scale data sets. We also find that from the first data set to the fifth data set, the difference in time consuming between the two algorithms is greatly obvious, but in Led17 data set and in SkillCraft1 Master Table Data set, the time consuming of the proposed discernibility matrix approach is almost the same as the time consuming of the granular space reduction based on Heuristic approach. With the increase of the size of the data set, the difference of the time consuming between the two algorithms is smaller and smaller. This phenomenon show that the proposed discernibility matrix approach is more suitable for small-scale data set, and the granular space reduction based on Heuristic approach is more suitable for large-scale data set.



7 Conclusions

To access some brief and effective decision rules with a threshold from incomplete information decision system under multigranulation environment, good and convenient granular space reduction is needed. This paper has introduced four kinds of the granular space reduction called a tolerance relations optimistic/pessimistic multigranulation β lower/upper approximation distribution reduction, which preserve the optimistic/pessimistic multigranulation β lower/upper approximation distribution of the decision classes. The judgement theorems and discernibility matrices associated with the four reductions have been obtained. Finally, an illustrative example has been applied to explain the mechanism of the proposed method. The proposed discernibility matrix approach renders a set of simpler discernibility functions for finding all approximation reductions of an incomplete decision system under multigranulation environment. Through

experiments, the advantage of the proposed method get a better reflection.

References

[1] Pawlak Z 1982 Rough Set *International Journal of Computer and Information Sciences* **11**(5) 341-56
 [2] Ananthanarayana V S, Narasimha M M, Subramanian D L 2003 Tree structure for efficient data mining using rough sets *Pattern Recogn.Lett* **24** 851-62
 [3] Chen Y L, Weng C H 2009 Mining fuzzy association rules from questionnaire data *Knowl.-Based Syst* **22** 46-56
 [4] Hulse J V, Khoshgoftaar T 2009 Knowledge discovery from imbalanced and noisy data *Data Knowl.Eng* **68** 1513-42
 [5] Zhai L Y, Khoo L P, Zhong Z W 2009 A dominance-based rough set approach to Kansei Engineering in product development *Expert Syst Appl.* **36** 393-402
 [6] Li G D, Yamaguchi D, Nagai M 2008 A grey-based rough decision-making approach to supplier selection *Int.J.Adv.Manuf.Techn* **36** 1032-40
 [7] Liou J J H 2009 A novel decision rules approach for customer relationship management of the airline market *Expert Syst Appl.* **36** 4374-81
 [8] Xie G, Zhang J, Lai K K, Yu L 2008 Variable precision rough set for group decision-making:an application *Int.J.Approx. Reason* **49** 331-43
 [9] Parthala N M, Shen Q 2009 Exploring the boundary region of tolerance rough sets for feature selection *Pattern Recognition* **42** 655-67
 [10] Wang J Y, Zhou J 2009 Research of reduct features in the variable precision rough set model *Neurocomput* **72** 2643-48
 [11] Qian Y H, Liang J Y, Yao Y Y, Dang C Y 2012 MGRS:A multi-granulation rough set *Inform.Sci.* **180** 949-70
 [12] Qian Y H, Liang J Y, Yao Y Y, Dang C Y 2010 Incomplete multigranulation rough set *IEEE T.Syst.Man Cy.B.* **20** 420-31
 [13] Raghavan R, Tripathy B K 2011 On some topological properties of multigranular rough sets *Advances in Applied Science Research* **2**(3) 536-43
 [14] Xu W H, Zhang X T, Wang Q R 2012 A generalized multi-granulation rough set approach *Lecture Notes in Computer Sciences* **6840** 681-9
 [15] Xu W H, Zhang X T, Wang Q R 2011 Multi-granulation fuzzy rough sets in a fuzzy tolerance approximation space *International Journal of Fuzzy Systems* **13**(4) 246-59
 [16] Xu W H, Zhang X T, Wang Q R 2012 Multiple granulation rough set approach to ordered information systems *International Journal of General Systems* **41**(5) 475-501
 [17] Ziarko W 1993 Variable precision rough set model *Journal of Computer System Science* **46**(1) 39-59
 [18] Skowron A, Rauszer C 1992 The Discernibility Matrices and Functions in Information systems *Kluwer Academic Publishers*
 [19] Kryszkiwicz M 1998 Rough set approach to incomplete information systems *Informa.Sci.* **112** 39-49
 [20] Zhongmei Zhou, Xuejun Wang, Guiying Pan 2012 CMCCR: Classification Based on Multiple Class-Correlation Rules *Journal of Digital Information Management* **10**(2) 64-70
 [21] Vairis A 2012 Mathematical Modelling of the Linear Friction Welding Process *Journal of Engineering Science and Technology Review* **5**(3) 6-9

Authors	
	<p>Sheng Yao, born in October, 1979, Hefei City, Anhui Province, P.R. China</p> <p>Current position, grades: the lecturer of Faculty of Computer science and Technology, An hui University. University studies: master candidate of Faculty of Computer science and Technology, An hui University. Scientific interest: Rough set, data mining. Publications: more than 10 papers published in various journals. Experience: Graduated from An hui University in 2002, more than 10 papers published in various journals.</p>
	<p>Longshu Li, born in February, 1956, Hefei City, Anhui Province, P.R. China</p> <p>Current position, grades: professor, doctoral supervisor of Faculty of Computer science and Technology, An hui University. University studies: Graduated from Anhui University of Computer science and Technology in 1996, received a master's degree in Computers. Scientific interest: Rough set. Machine Learning, Data mining. Publications: Presided over the completion of national and provincial, than 10 scientific research projects; more than 80 papers published in various journals. Experience: Graduated from Anhui University of Computer science and Technology in 1996, received a master's degree in computers, was approved as a tutor of doctoral students in 2006; has completed including the national and provincial and ministerial level, than 10 scientific research projects; more than 80 papers published in various journals published monographs, more than 5 teaching materials.</p>

Cost-sensitive back-propagation neural network for financial distress prediction

Wang Hongbao*

College of Applied Foreign Language, Heilongjiang University, No. 74 Xuefu Road, Harbin, Heilongjiang Province, China, 150080

Received 27 January 2014, www.tsi.lv

Abstract

Financial distress prediction (FDP) models, which classify financially distressed companies from healthy ones, prevent market participants from suffering economic loss. In the process of FDP, the misclassification of type I error of the model incurs much higher cost than that of type II error. Most of the previous FDP models do not take the asymmetric costs into consideration. In this paper, cost-sensitive back-propagation neural network (CS-BPNN) FDP model is proposed for minimizing the cost of prediction error such that the loss of users of the model will suffer less. The performance of the model is evaluated by taking 180 Chinese listed companies as sample data and adopting 8 times of sampling to assess different misclassification costs and prediction accuracy. The experimental results suggest that the proposed approach helps to improve the prediction performance in asymmetric cost setup.

Keywords: Financial Distress Prediction (FDP), Cost Sensitive Learning, Back-propagation Neural Network (BPNN), Cost of Prediction Error, Cost-sensitive Back-propagation Neural Network (CS-BPNN)

1 Introduction

As the internal and external environment of a company changes in speed and complexity, business organization with management deficiency and lack of innovation may be very likely to lead to financial distress and even bankruptcy. Financial distress prediction (FDP) has a key influence on the enterprise's development and its stakeholders' decision as well [1]. For a commercial bank, FDP has profound impact on its credit scoring because banks should watch the current and future financial status of their enterprise customers all the time. For shareholders, FDP facilitates to detect the financially distressed condition of a company in advance so that they will withdraw capital before suffering huge economic loss [2]. Therefore, FDP has been a major research area within corporate finance for decades.

The earliest popular techniques in FDP were the statistical models, such as univariate analysis [3], multivariate discriminant analysis (MDA) [4] and logistic regression (Logit) and etc. [5]. Since 1990s, artificial intelligent and data mining techniques took a key role in FDP with the rapid development of computer technology [6]. Neural network (NN) has become one of the most widely used machine learning techniques in FDP due to its strong nonlinear mapping ability. Many researchers compared NN models with MDA and Logit and concluded that the prediction accuracy of NN models was higher than MDA and Logit [7]. Apart from NN, other artificial intelligent techniques were also employed in FDP, such as decision tree [8], genetic algorithm [9], rough sets and etc. [10]. Support vector machine (SVM)

is a relatively new machine learning technique and is widely applied in many fields, such as classification, data mining and time series forecasting [11]. SVM is superior to other algorithms for FDP in situations where the variables demonstrate complex nonlinear relationships. However, it still has problems with identifying the relative importance of variables and searching the optimal parameters [12].

These previous classification techniques aim to minimize overall error based on the consumption that the misclassification costs of type I error and type II error are equal. However, this assumption is not valid in FDP, where the cost of misclassifying a distressed company as a healthy one is much higher than the inverse. Therefore, the asymmetric cost information should be taken into consideration in FDP so that different stakeholders could select their favourable models based on their cost preference. Ref [13] incorporated cost information into learning vector quantization (LVQ) approaches in FDP. To our knowledge, the study of cost-sensitive back-propagation neural network (CS-BPNN) is not reported for FDP.

As a result, the main motivation of this paper is to employ CS-BPNN to establish companies' FDP model in order to minimize the cost of prediction error. The main objectives of this paper are to (1) incorporate cost information into BPNN algorithm to make the traditional BPNN cost-sensitive in FDP, (2) exclude the missing and outlier data in the initial data pre-processing stage and use statistical methods to screen financial ratios in order to improve the prediction accuracy of FDP model, (3) compare CS-BPNN approach with BPNN approach in the

* Corresponding author- Tel: +86-139-3650-1851; fax: +86-139-3650-1851; E-mail: Lodgewang@163.com

aspect of cost of misclassification and prediction accuracy, and (4) expand CS-BPNN approach so that it will provide decision makers with evidence in model selection.

2 Research Background

2.1 NEURAL NETWORK

Among different neural network architectures, BPNN is the most frequently employed architecture due to its simplicity and excellent performance in extracting useful information from samples [14]. P_R denotes the elementary inputs of BP, as shown in Fig. 1. W_i is an appropriate weight of each input. The sum of the weighted inputs and the bias is input to the f function, which transforms the sum of input value into output value of the node.

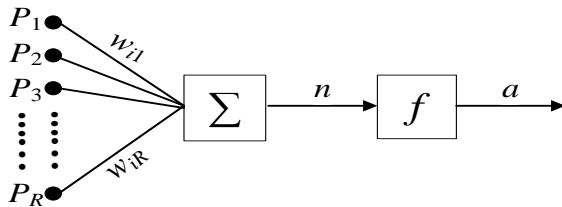


FIGURE 1 Artificial neuron model

2.2 COST-SENSITIVE BPNN

Cost-sensitive learning solves the problem in which different misclassification errors correspond with different costs. The aim of our paper is to make the misclassification cost minimum. If i equals to j , the misclassification cost of sample x classified as j is zero, shown as Equation 1:

$$C(i | x) = \sum_{j=1}^c C(i, j)P(j | x), \tag{1}$$

where c is the misclassification cost, i and j are the two classification class and x is a sample. If i does not equal to j , the misclassification cost of sample x classified as j is c .

In order to make c minimum, the value can be obtained by the following equation:

$$C(i|x) = \min_{i \in c} \arg \sum_{i \in I} C(i, j)P(j|x), \tag{2}$$

The purpose of training BPNN is to make mean square deviation least:

$$f(x) = E[(t - a)^2] = \sum_{i=1}^N p_i(t_i - a_i)^2, \tag{3}$$

Cost-sensitive BPNN takes the misclassification minimum as the evaluation indicator of a model performance, as shown in Equation 4:

$$F(x) = E[C(i, j)] = \sum p(i)p(j|i)C(j, i), \tag{4}$$

In FDP, the output layer has only one neuron, and the misclassification cost of the sample x is shown in the Equation 5:

$$\mu(x) = \frac{1}{2}(t_k - a_k)^2 C(t_k, |a_k|), \tag{5}$$

3 Empirical Experiment

3.1 EXPERIMENT DATA

3.1.1 Initial data collection and pre-processing

Because the empirical research is carried out on real world information of Chinese listed companies, financially distressed companies are defined as those who have had negative net profit in consecutive two years, or its net capital per share is lower than the face value per share for the reason of one year's substantive loss. Therefore, the financially distressed companies are specially treated (ST) by Chinese Stock Exchange. Healthy companies are defined as those who have never been specially treated.

Since the financial ratios of ST companies have been deteriorating even before ST, the adoption of financial data one year before ST results in overestimation of the prediction performance of the model. In this paper, the financial data from two years to five years before ST is selected, namely $U_{(t-2)}$, $U_{(t-3)}$, $U_{(t-4)}$ and $U_{(t-5)}$. The data used in this research is obtained from RESSET Financial Database. 90 pairs of companies listed in Shenzhen Stock Exchange and Shanghai Stock Exchange are selected as initial data set. 30 financial ratios are selected as initial features, covering debt ability, growth ability, capital structure, activity ability, profitability and indicators per share. In order to eliminate outlier data and missing data, companies with financial ratios deviating from the mean value as much as four times of standard deviation are excluded and companies missing at least one financial ratio are also excluded. The final number of sample data is 170.

3.1.2 Experimental data sets

The empirical experiment aims to validate whether FDP model based on CS-BPNN can minimize the cost of prediction error. 57 pairs of financially distressed companies and healthy companies are selected to form training data set and the rest 28 pairs are used to form testing data set.

3.1.3 Feature selection

In the field of FDP, a large number of financial indicators are usually involved in order to obtain an accurate financial condition of companies. However, some financial indicators cannot precisely identify financially distressed companies from healthy ones. Therefore, the purpose of feature selection addresses the problem by removing irrelevant and redundant features, improving the accuracy of the model, decreasing the computational effort and facilitating the use of the model.

(1) Statistical analysis

In this paper, statistical methods are used to screen financial ratios. The selection procedure is as follows: Kolmogrov-Simironov test is employed to examine whether each financial ratio meets normal distribution. If financial ratios meet normal distribution, T test is employed to validate whether the financial ratios are significant. If financial ratios do not meet normal distribution, Mann-Whitney test is employed to validate whether the financial ratios are significant, as shown in Table 1.

TABLE 1 Significance test of financial ratios in year $t-2$

Variables	T test		Kolmogrov-Simironov test		Mann-Whitney test	
	T-Statistic	Prob.	KS-Statistic	Prob.	MW-Statistic	Prob.
Return on Equity			5.365	**0.000	1263.000	**0.000
Return on Assets			2.323	**0.000	2096.000	**0.000
Return on Invested Capital			3.101	**0.002	1744.000	**0.000
Net Profit Margin			6.259	**0.000	1874.000	**0.000
Cost Profit Margin			2.752	**0.000	1797.000	**0.000
Current Ratio			2.420	**0.000	1086.000	**0.000
Quick Ratio			2.862	**0.000	1233.000	**0.000
Equity Ratio			5.055	**0.000	1120.000	**0.000
Debt to Asset Ratio			5.226	**0.000	1210.000	**0.000
Debt to Tangible Asset Ratio			4.602	**0.005	3287.000	0.310
Operating Cash Flow/Total Liability			6.785	**0.002	2146.000	**0.000
Operating Income Growth Rate			5.614	**0.000	2786.000	*0.010
Net Profit Growth Rate			4.696	**0.000	3072.000	0.092
Total Asset Growth Rate			4.209	**0.001	1495.000	**0.000
Turnover Rate of Accounts Receivable			5.532	**0.000	3508.000	0.745
Turnover Rate of Accounts Payable			2.723	**0.000	2435.000	**0.000
Turnover Rate of Current Assets			1.872	**0.002	3392.000	0.492
Turnover Rate of Fixed Assets			5.852	**0.000	2003.000	**0.000
Turnover Rate of Equity			4.419	**0.000	2637.000	**0.002
Turnover Rate of Total Assets			1.559	*0.015	2088.500	**0.000
Earning Per Share			2.708	**0.000	1578.500	**0.000
Net Asset Value Per Share	7.666	**0.000	1.181	0.123		
Operating Revenue Per Share			2.355	**0.000	1114.000	**0.000
Gross Profit Margin			1.529	*0.019	2485.000	**0.000
Net Return on Assets			3.634	**0.000	1868.000	**0.000
Fixed Assets Ratio	-2.760	**0.007	1.075	0.198		
Equity Ratio			5.226	**0.000	1210.000	**0.000
Operating Profit Growth			5.807	**0.000	2898.000	**0.026
Earning Per Share Growth			4.611	**0.000	3114.500	0.121
Every Dividend Profit before Tax			2.604	**0.000	1698.000	**0.000

Note: *Significant at 5%; **Significant at 1%.

(2) Analysis on significance test of financial ratios

As shown in table 1, only net asset value per share and fixed assets ratio pass Kolmogrov-Simironov test. The result is consistent with the previous research conclusion that most financial ratios do not meet normal distribution. Additionally debt to tangible asset ratio, operating cash flow/total liability, net profit growth rate, total asset growth rate, turnover rate of accounts receivable, turnover rate of current assets, turnover rate of equity, operating revenue per share, gross profit margin, net return on assets, fixed assets ratio, equity ratio, operating profit growth, earning per share growth and every dividend profit before tax do not pass the significance test from year $t-2$ to year $t-5$. Therefore, these 15 ratios are discarded, by which healthy companies cannot be distinguished from distressed companies.

3.2 PARAMETER SETTING

3.2.1 Setting of cost of misclassification

When misclassification occurs in FDP model, the costs of misclassification of different stakeholders are different. For example, company shareholders' loss is hugely different from managers' one when FDP model has type I error. Therefore, the user of the FDP model should be identified in the first place. Then cost matrix is used to determine the value of cost.

In the empirical research, the user of FDP model is supposed to be a commercial bank. The bank can use the prediction result of FDP model to make a decision of making loans or not. Therefore, when the model has type

I error, the cost of error of the bank is the full loan. When the model has type II error, the cost is the loan interest loss. Since loan interest rates of a commercial bank to an enterprise in China range from 5% to 30% above personal loan interest rate, 30% above personal loan interest rate is taken into calculation for the sake of unified computing.

3.2.2 Setting of parameters of CS-BPNN

Classification performance of CS-BPNN is affected mainly by the parameter pair, learning rate α and number of training round g , whose optimal parameter combination is determined by grid method and leave-one-

out cross validation test. The combination, which has the least misclassification cost in the training set data, is selected as the parameters of CS-BPNN, where $\alpha \in \{10^{-3}, 10^{-2}, 10^{-1}, 3 \times 10^{-1}\}$ and $g \in \{100, 500, 1000, 3000, 5000, 10000\}$. As shown in table 2-5, the optimal parameter value in year $t-2$ is $\alpha_{(t-2)} = 0.1, g_{(t-2)} = 10000$. The optimal parameter value in year $t-3$ is $\alpha_{(t-3)} = 0.3, g_{(t-3)} = 5000$. The optimal parameter value in year $t-4$ is $\alpha_{(t-4)} = 0.3, g_{(t-4)} = 5000$. The optimal parameter value in year $t-5$ is $\alpha_{(t-5)} = 0.3, g_{(t-5)} = 3000$.

TABLE 2 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-2

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
a=0.001	Type I error	0	0	0	0	0	0
	Type II error	60	60	60	60	60	59
	Total error	5.4	5.4	5.4	5.4	5.4	4.86
a=0.01	Type I error	0	0	0	2	3	6
	Type II error	60	60	59	47	41	36
	Total error	5.4	5.4	4.86	6.23	6.69	9.24
a=0.1	Type I error	0	2	3	3	3	2
	Type II error	59	43	38	28	24	20
	Total error	4.86	5.87	6.42	5.52	5.16	3.8
a=0.3	Type I error	2	4	3	3	2	3
	Type II error	49	34	34	20	36	22
	Total error	6.41	7.06	6.06	4.8	5.24	4.98

TABLE 3 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-3

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
a=0.001	Type I error	0	0	0	0	0	0
	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
a=0.01	Type I error	0	0	0	2	3	4
	Type II error	60	60	60	50	44	37
	Total error	5.4	5.4	5.4	6.5	6.96	7.33
a=0.1	Type I error	0	3	4	3	2	4
	Type II error	60	44	38	35	34	25
	Total error	5.4	6.96	7.42	6.15	5.06	6.25
a=0.3	Type I error	2	3	3	2	2	3
	Type II error	51	39	36	35	25	25
	Total error	6.59	6.51	6.24	5.15	4.25	5.25

TABLE 4 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-4

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
a=0.001	Type I error	0	0	0	0	0	0
	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
a=0.01	Type I error	0	0	0	1	2	3
	Type II error	60	60	60	55	47	45
	Total error	5.4	5.4	5.4	5.95	6.23	7.05
a=0.1	Type I error	0	2	3	2	4	4
	Type II error	60	47	46	43	38	32
	Total error	5.4	6.23	7.14	5.87	7.42	6.88
a=0.3	Type I error	1	1	3	2	3	4
	Type II error	45	45	43	33	35	32
	Total error	5.05	5.05	6.87	4.97	6.15	6.88

TABLE 5 Leave-one-out cross validation test result of cost-sensitive BP neural network in year t-5

		g=100	g=500	g=1000	g=3000	g=5000	g=10000
a=0.001	Type I error	0	0	0	0	0	0
	Type II error	60	60	60	60	60	60
	Total error	5.4	5.4	5.4	5.4	5.4	5.4
a=0.01	Type I error	0	0	0	2	4	5
	Type II error	60	60	60	59	57	50
	Total error	5.4	5.4	5.4	7.31	9.13	9.5
a=0.1	Type I error	0	4	4	3	4	5
	Type II error	60	57	53	36	31	28
	Total error	5.4	9.13	8.77	6.24	6.79	7.52
a=0.3	Type I error	2	3	2	2	5	2
	Type II error	59	46	40	33	34	34
	Total error	7.31	7.14	5.6	4.97	8.06	5.06

3.3 EXPERIMENT RESULTS AND ANALYSIS

In order to verify the prediction performance of FDP model based on CS-BPNN, the empirical research makes a comparison with the one based on BPNN. The optimal parameter combination of FDP model based on BPNN is also determined by grid method and leave-one-out cross validation test. In order to get multiple performance

statistics, multiple experimental data sets are formed by repetitively and randomly classifying training sample and testing sample. By 8 times of random sampling without replacement, 57 pairs of financial distressed companies and healthy ones are selected as training data set and the rest 28 pairs are selected as testing data set each time, as shown in table 6.

TABLE 6 Experimental results on testing data set in year t-2

Data	BPNN				CS-BPNN			
	Number of type I error	Number of type II error	Prediction accuracy	Total cost of prediction error	Number of type I error	Number of type II error	Prediction accuracy	Total cost of prediction error
set 1	4	5	83.93	4.45	2	8	82.14	2.72
set 2	1	2	94.64	1.18	1	2	94.64	1.18
set 3	6	4	82.14	6.36	2	9	80.36	2.81
set 4	4	4	85.71	4.36	4	9	76.79	4.81
set 5	4	1	91.07	4.09	2	3	91.07	2.27
set 6	2	1	94.64	2.09	1	2	94.64	1.18
set 7	8	4	78.57	8.36	4	9	76.79	4.81
set 8	6	3	83.93	6.27	4	4	85.71	4.36
Average	4.38	3	86.83	4.65	2.5	5.75	85.27	3.02

TABLE 7 Comparison of test results of BPNN and CS-BPNN from year t-2 to year t-5

Year	FDP models	BPNN	CS-BPNN	Year	FDP models	BPNN	CS-BPNN
t-2	Average Number of type I error	4.38	2.5	t-3	Average Number of type I error	5.31	2.1
	Average Number of type II error	3	5.75		Average Number of type II error	2.45	8.6
	Total cost	4.65	3.02		Total cost	5.53	2.87
t-4	Average Number of type I error	7	4.2	t-5	Average Number of type I error	6.12	2
	Average Number of type II error	4.33	9.36		Average Number of type II error	10.65	15.5
	Total cost	7.39	5.04		Total cost	7.08	3.4

As shown in table 6, the average prediction accuracy of FDP model based on BPNN is 86.83, which is slightly higher than that of FDP model based on CS-BPNN, 85.27. It is mainly because FDP model based on CS-BPNN integrates the different costs of prediction errors. Since the cost of type I error is hugely larger than that of type II error, the average cost of prediction error of FDP model based on BPNN is 4.65, which is much higher than that based on CS-BPNN, 3.02.

In table 7, the empirical results suggest that the prediction performance of both BPNN and CS-BPNN become weaker with the selection of earlier training data sets. However, we do not draw the same conclusion in the misclassification cost. For example, the misclassification cost of CS-BPNN in year t-5 is lower than in year t-4. The main reason is that even though the prediction accuracy of CS-BPNN in year t-4 is higher

than t-5, the number of type I error in year t-4 is more than the one in year t-5.

The experimental results and analysis suggest that BPNN achieves slightly better prediction accuracy than CS-BPNN. However, CS-BPNN produces a much better result than BPNN in the total misclassification costs. With the selection of earlier training data sets, BPNN and CS-BPNN become weaker in the prediction accuracy, but they do not perform the same way in the misclassification cost.

4 Conclusion

Financial distress prediction is extensively studied in the corporate governance field. Few studies incorporate unequal misclassification costs into FDP model. Cost-sensitive classification models, coping with asymmetric

costs of type I error and type II error, are of crucial interest to stakeholders' decisions. This paper verifies how the asymmetric costs of two kinds of errors are integrated into FDP model. This research takes 85 financial healthy companies and matches them with 85 financially distressed companies. 114 companies are selected as training data set and the rest are selected as testing data set. Experimental tests demonstrate that CS-BPNN approach leads to a lower total misclassification cost when compared with the traditional BPNN one.

Acknowledgments

In the process of writing paper, thanks so much for the help and understanding of my wife and my daughter.

References

- [1] Zhi X, Xianglei Y, Ying P, Xin D 2012 The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory *Knowledge-Based Systems* **26** 196-206
- [2] Pochang K, Pingchen L 2006 An evolution-based approach with modularized evaluations to forecast financial distress *Knowledge-Based Systems* **19**(1) 84-91
- [3] Fitzpatrick P J 1932 A comparison of ratios of successful industrial enterprises with those of failed firms *Certified Public Accountant* **10**(2) 598-605
- [4] Altman E I 1968 Financial ratios discriminant analysis and the prediction of corporate bankruptcy *Journal of Finance* **23**(4) 589-609
- [5] Ohlson J A 1980 Financial ratios and the probabilistic prediction of bankruptcy *Journal of Accounting Research* **18**(1) 109-31
- [6] Chaudhuri A, De K 2011 Fuzzy support vector machine for bankruptcy prediction *Applied Soft Computing* **11**(2) 2472-86
- [7] Odom M, Sharda R 1990 A neural networks model for bankruptcy prediction *IEEE International Conference on Neural Network* (San Diego) **2** 163-8
- [8] Frydman H, Altman E I, Kao D 1985 Introducing recursive partitioning for financial classification: the case of financial distress *Journal of Finance* **40**(1) 269-91
- [9] Lensberg T, Eilifsen A, McKee T E 2006 Bankruptcy theory development and classification via genetic programming *European Journal of Operational Research* **169**(2) 677-97
- [10] McKee T E 2000 Developing a bankruptcy prediction model via rough sets theory *International Journal of Intelligent Systems in Accounting, Finance and Management* **9**(3) 159-73
- [11] Haibin, X, Muhua L, Haichao Y, Qian H, Jinhui Z 2013 Study on detection and classification of Tetracycline residue in duck meat using synchronous fluorescence spectra and support vector machine *Journal of Engineering Science and Technology Review* **6**(2) 85-9
- [12] Vapnik V N 1995 *The Nature of Statistical Learning Theory* Springer-Verlag: New York
- [13] Ning C, Bernardete R, Armando S V, Joao D, Joao C N 2010 Hybrid genetic algorithm and learning vector quantization modelling for cost-sensitive bankruptcy prediction *2nd International Conference on Machine Learning and Computing* (Bangalore) 213-7
- [14] Shujiang F, Yi Z, Honghua X, Chuang L 2012 Risk evaluation of rural financial organization operation based on the BP *Journal of Digital Information Management* **10**(5) 341-6

Author



Hongbao Wang, born in April, 1979, Harbin City, Heilongjiang Province, P.R. China

Current position, grades: Lecturer, College of Applied Foreign Language, Heilongjiang University, Harbin, China.

University studies: received his M.Sc. in Investment Analysis from Stirling University in UK. He received his Ph.D. in Corporate Governance from Harbin Institute of Technology in China.

Scientific interest: His research interest fields include company financial distress prediction and data mining.

Publications: more than 10 papers published in various journals.

Experience: He has been teaching for 9 years and has completed three scientific research projects.

Fuzzy location-routing problem for emergency logistics systems

Shaoren Wang^{1*}, Zujun Ma², Bochao Zhuang¹

¹ College of Business Administration, HuaQiao University, Quanzhou City, Fujian Province, 362021, China

² School of Economics and Management, Southwest Jiaotong University, Chengdu City, Sichuan Province, 610031, China

Received 1 March 2014, www.tsi.lv

Abstract

To optimize the post-earthquake emergency logistics system with two-echelon multi-facilities, this study develops a model for fuzzy location-routing problem by considering fuzzy demand of relief materials, timeliness and limited resources. The goal of the model is to minimize the total cost and the relief time of system. Furthermore, this research proposes an improved genetic algorithm based on weighted coefficient transformation. The result of a numerical example shows that the model and algorithm are effective for resolving the joint decision-making of facility location-allocation problem and vehicle routing problem in post-earthquake.

Keywords: Emergency Logistics, Location-routing Problem, Earthquake Disasters, Multi-objective Optimization, Improved Genetic Algorithm

1 Introduction

According to the statistical findings of Sichuan Red Cross, after the world-shocked earthquake happened in Wenchuan, people in the harder hit area demand 4436.2 tons of food, 8115 tons of drinking water every day. The category of relief supplies reached 7 categories and 46 subcategories, which became a great challenge to relevant dispatching department. Due to lack of scientific and reasonable planning emergency logistics system scheme, some of the disaster areas were repeated distributed which resulted in excess supplies, while some harder hit area had not received emergency relief supplies which led to low efficiency of emergency logistics. We can conclude from the actual feedback information of the earthquake relief work, that it is urgent to solve the following problems in the optimization of post-earthquake emergency logistics system: How to locate the emergency facilities? How to optimize the transport routes of emergency supplies? The first question is to solve the emergency facility location-allocation problem (Location-Allocation Problem, LAP). It determines the mode and structure of the whole emergency logistics system. While the second question is mainly to solve the problem of emergency vehicle routing problem (Vehicle Routing Problem, VRP). Both of them are key issues in optimizing post-earthquake emergency logistics system, they are interdependent and interactive. Therefore, from the perspective of optimizing the overall emergency logistics system, investigating the scheme of location-allocation and route problem (LRP) in the earthquake emergency logistics system is quite necessary.

Although many scholars (Min H [1], Nagy G [2], Sérgio Barreto [3]) have studied the LRP in the general logistics systems, most of research (Haghani A [4],

Barbarosoğlu G [5], WEI Hang [6], Yuan Yuan [7], Sheu J B [8, 9]) on the optimization of emergency logistics system basically focuses on emergency facility location problem or emergency supplies distribution alone. However, Research on LRP in emergency logistics system is rare. As Yi et al [10] explore the coordination and optimization of materials distribution and treatment and the transportation of the wounded in the natural disaster emergency rescue. They set up a deterministic mixed multiple integer and multispecies goods network flow model, and use CPLEX software to obtain the solution. In the emergency response phase, quickly drawing out the emergency rescue plan is the key of improving the efficiency. Some scholars make some contributions in this stream research, for example, Wang et al [11], designed an "triangle" heuristic algorithm based on the two-stage decomposed thoughts in a two-echelon post-earthquake facility location routing problem (LRP) model. They prove that "triangle" heuristic algorithm has higher efficiency by comparing it with the improved genetic algorithm. Nevertheless, their study does not consider the fuzziness of post-earthquake emergency material demand. In fact, the information of relief materials' supply and demand, traffic, the emergency resource and other factors is often uncertain. It is a fuzzy decision process. Besides, natural disaster emergency logistics shows particularity characteristics, such as relief materials demand can be divided, and each material requirement can be accessed multiple times. Full load direct distribution and touring (circuit) distribution methods can be used alternately through analysing the different categories of disaster relief materials. The demand and supply of natural disaster emergency materials have great uncertainty, it is necessary to coordinate the relief materials supply and the inaccurate

* Corresponding author - Tel: +86-595-226-90276; fax: +86-595-226-90276; E-mail: okwangshao@163.com

prediction of demand [12]. Earthquake emergency logistics has the typical features of sudden, uncertainty and timeliness, therefore, this study aims at optimizing the post-earthquake emergency logistics system with two-echelon multi-facilities, setting up an optimized multi objective model for fuzzy location-routing problem (LRP), proposing an improved genetic algorithm based on weighted coefficient transformation, and to test the improved genetic algorithm from the standpoint of its robustness and its parameters setting.

2 Model construction

2.1 PROBLEM DESCRIPTION

After the earthquake, it is necessary to establish a number of relief materials collector-distributor points in the periphery to conduct the transfer and allocation of all donations and emergency procurements. It can help to quickly deliver the materials to the temporary emergency distribution center and then send them to the disaster area settlements (shelters). Because of the fuzzy demand of relief materials, the last visited settlements (shelters) of each route may appear unmet demand. It needs to dispatch helicopters from the airport to the existing relief materials distribution point to load supplies, and then do urgent touring (circuit) distribution to satisfy these settlements (shelters)' demands. The decision problem is how to establish several relief materials collector-distributor points and emergency distribution centres, and how to determine a set of emergency transportation routes after considering the facility capacity and rescue vehicles, helicopters capacity constraints, meanwhile, to minimize the operation time of emergency logistics system and the total cost. Assumptions before modelling show as follows:

- (1) This study only consider about the distribution of some kind of relief materials.
- (2) No material transport between the relief materials collector-distributor points.
- (3) Assuming the vehicle capacity of the collector-distributor points and emergency distribution centres can meet the transportation demands.
- (4) Large demand settlements (shelters) can be accessed several times, use some idle vehicles with full load direct distribution, the remaining less than a car of materials with touring (circuited) distribution.
- (5) Since the demand for each settlement (shelter) is fuzzy, it can be represented by triangular fuzzy number.
- (6) Helicopters can be dispatched urgently from the airport to the existing collector-distributor points to load materials for touring (circuited) distribution, when the settlement (shelter) of each route appears unmet needs.

2.2 SYMBOL ILLUSTRATION

$S = \{r | r = 1, 2, 3, \dots, R\}$ indicates alternative collector-distributor points set; SQ_r indicates the maximum management capacity of distribution r (capacity).

$W = \{p | p = 1, 2, 3, \dots, P\}$ indicates alternative distribution centers set; WQ_p indicates maximum management capacity of center p (capacity).

$O = \{a | a = 1\}$ indicates helicopter airports set, the airports locations are determined.

$C = BC \cup SC$ indicates the disaster settlements (shelters) set. BC indicates settlement (shelter) with large demand (Demand exceeds the maximum capacity of the vehicle) (Large Shelter); SC indicates settlement (shelter) with small demand (Demand is less than the maximum loading capacity of the vehicle) (Small Shelter).

$N = S \cup W \cup O \cup C$ is the set of all the nodes.

$V = \{k | k = 1, 2, 3, \dots, K\}$ is the set of vehicles.

VQ_k indicates the loading capacity of vehicle k ; VQ indicates the maximum loading capacity of the largest volume vehicle.

$U = \{h | h = 1, 2, \dots, H\}$ indicates the helicopters set; UQ_h indicates the capacity of helicopter h .

DT indicates the time cost by the full loading vehicle with per unit of material and distance from distribution centre to the big shelter.

CV_k stands for the dispatch money of vehicle k .

CU_h stands for the dispatch money of helicopter h .

CB stands for the transportation expenses of per unit of material and distance from collector-distributor point to emergency distribution centre (or from distribution centre to the big shelter).

D stands for the transportation expenses of (tour) circuited vehicle with certain unit of distance.

A stands for the transportation expenses of (tour) circuited helicopter with certain unit of distance.

CR_r stands for the cost of establishing r collector-distributor point, assuming that it is positive associated with the scale of the distribution point.

CW_p stands for the cost of establishing distribution center at point p , assuming that it is positive associated with the scale of the distribution centre.

β_{ij} stands for the road damage rate between node i and node j ($i, j \in N$), represented by the ration of the length of damage road to the length of the whole road.

η , γ separately stands for the average cost and average time spent on road repairing of per unit distance between node i and node j ($i, j \in N$).

$\alpha_{ij} \in \{1, 0\}$ stands for the road situation, 1 indicates that the road between node i and j ($i, j \in N$) is connected, 0 indicates that the road is blocked. β_{ij} , η , α_{ij} , γ can be identified by the remote sensing data of satellite, aerial

technology, some practical road feedback information and calculation of the amount of construction.

$$\tilde{d}_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \text{indicates the}$$

Euclidean distance between node i and node j ($i, j \in N$), x_i, y_i ($i \in N$) indicates the abscissa and ordinate of node i .

d_{ij} indicates the actual distance between node i and node j ($i, j \in N$). This can be obtained by GIS calculation, here we replace it with the Euclidean distance transformation, that is $d_{ij} = m(1 - \alpha_{ij}) + \delta \tilde{d}_{ij}$, δ represents the coefficient of the plane Euclidean distance transforms into actual distance ($\delta > 1$), the value can be estimated according to the disaster area road network information; m is a sufficiently large positive number.

q_j indicates the demand of shelter j ($j \in C$), which can be represented by Triangular fuzzy number (a_j, b_j, c_j) , a_j, b_j, c_j can be estimated by the decision makers based on intuition, experience or some available information.

\bar{q}_j indicates the unmet demand of the last visited settlements (shelters) j ($j \in C$) of each route, $\bar{q}_j \geq 0$, it is a definite number, for if the surplus materials on the car can meet the needs of the settlements (shelters) can be known once the car reach the final service settlements (shelters).

ST_{jk} indicates the time that vehicle k spends on arriving at the small settlement (shelter) j ($j \in SC$), when $j \in W$, $ST_{jk} = 0$.

v_k indicates the average speed of the No. k car.

τ_{ijk} indicates the time that vehicle costs from node i to node j ($i, j \in W \cup SC$), $\tau_{ijk} = d_{ij} / v_k$.

BT_j indicates the traffic time of sending full load vehicle to the big settlement (shelter) j ($j \in BC$).

ET_{jh} indicates the traffic time of helicopter h reaching the demand-unmet small settlement (shelter) j ($j \in SC$), when $j \in S \cup O$, $ET_{jh} = 0$.

v'_h indicates the average speed of the No. h helicopter.

ζ_{ijh} indicates the time that helicopter h costs from node i to node j ($i, j \in O \cup S \cup SC$), $\zeta_{ijh} = \tilde{d}_{ij} / v'_h$.

\tilde{p} indicates the confidence level of vehicle capacity limitation, $\tilde{p} \in [0, 1]$.

\bar{p} indicates the confidence level of capacity limitation of emergency distribution centre, $\bar{p} \in [0, 1]$.

The decision variables:

x_{ijk} : while vehicle k ($k \in V$) transports from node i to node j ($i, j \in W \cup SC$), it's 1, otherwise is 0.

φ_{ijh} : While helicopter h ($h \in U$) transports from node i to node j ($i, j \in O \cup S \cup SC$), it's 1, otherwise is 0.

u_r : while establishing distribution point at alternative point r ($r \in S$), it's 1, otherwise is 0.

z_p : while establishing emergency distribution centre at alternative point p ($p \in W$), it's 1, otherwise is 0.

g_{pr} : while established emergency distribution centre p ($p \in W$) is assigned to established collector-distributor point r ($r \in S$), it's 1, otherwise is 0.

y_{jp} : while established settlement (shelter) j ($j \in C$) is assigned to established emergency distribution centre p ($p \in W$), it's 1, otherwise is 0.

fs_{rp} indicates the transportation capacity from collector-distributor point r ($r \in S$) to emergency distribution centre p ($p \in W$).

fw_{pj} indicates the full load transportation capacity from emergency distribution centre p ($p \in W$) to the big settlement (shelter) j ($j \in BC$).

2.3 MODEL ESTABLISHING

Establishing fuzzy LRP multi-objective optimized model of post-earthquake emergency logistics system as follow:

$$\begin{aligned} \min f_1 = & \sum_{k \in V} \sum_{j \in SC} ST_{jk} + \sum_{j \in BC} BT_j + \sum_{h \in U} \sum_{j \in SC} ET_{jh} + \sum_{r \in S} \sum_{p \in W} \gamma \times \\ & \beta_{rp} \times d_{rp} \times g_{rp} + \sum_{p \in W} \sum_{j \in BC} \gamma \times \beta_{pj} \times d_{pj} \times y_{jp} + \quad , \quad (1) \\ & \sum_{k \in V} \sum_{i \in W \cup SC} \sum_{j \in W \cup SC} \gamma \times \beta_{ij} \times d_{ij} \times x_{ijk} \end{aligned}$$

$$\begin{aligned} \min f_2 = & \sum_{r \in S} CR_r \times u_r + \sum_{p \in W} CW_p \times z_p + \\ & \sum_{k \in V} CV_k \sum_{i \in W} \sum_{j \in SC} x_{ijk} + \sum_{h \in U} CU_h \sum_{i \in O} \sum_{j \in S} \varphi_{ijh} + \\ & \sum_{r \in S} \sum_{p \in W} (CB \times fs_{rp} + \eta \times \beta_{rp}) d_{rp} \times g_{rp} + \\ & \sum_{p \in W} \sum_{j \in BC} (CB \times fq_{pj} + \eta + \beta_{pj}) d_{pj} \times y_{jp} + \quad , \quad (2) \\ & \sum_{k \in V} \sum_{i \in W \cup SC} \sum_{j \in W \cup SC} (D + \eta \times \beta_{ij}) d_{ij} \times x_{ijk} + \\ & \sum_{h \in U} \sum_{i \in O \cup S \cup SC} \sum_{j \in O \cup S \cup SC} A \times \tilde{d}_{ij} \varphi_{ijh} \end{aligned}$$

$$s.t. \sum_{p \in W} fs_{rp} g_{pr} \leq SQ_r, \forall r \in S, \quad (3)$$

$$\sum_{r \in S} fs_{rp} g_{pr} \leq WQ_p, \forall p \in W, \quad (4)$$

$$\sum_{r \in S} fs_{rp} g_{pr} - \sum_{j \in C} q_j y_{jp} = 0, \forall p \in W, \quad (5)$$

$$\text{Pos}(\sum_{j \in C} q_j y_{jp} \leq WQ_p) \geq \bar{p}, \forall p \in W, \quad (6)$$

$$\text{Pos}(\sum_{j \in SC} q_j \sum_{i \in W \cup SC} x_{ijk} \leq VQ_k) \geq \tilde{p}, \forall k \in V, \quad (7)$$

$$\sum_{p \in W} f w_{pj} + \text{rep}(q_j, VQ) + \bar{q}_j \geq q_j, \forall j \in BC, \quad (8)$$

$$\varphi_{jh} = 1 \Rightarrow x_{jk} = 0, \forall j \in SC, i \in S \cup SC, l \in SC, h \in U, k \in V, \quad (9)$$

$$\sum_{j \in SC} \bar{q}_j \sum_{i \in O \cup S \cup SC} \varphi_{ijh} \leq UQ_h, \forall h \in U, \quad (10)$$

$$\sum_{i \in W \cup SC} x_{ijk} - \sum_{i \in W \cup SC} x_{jik} = 0, \forall j \in W \cup SC, k \in V, \quad (11)$$

$$\sum_{i \in O \cup S \cup SC} \varphi_{ijh} - \sum_{i \in O \cup S \cup SC} \varphi_{jih} = 0, \forall j \in O \cup S \cup SC, h \in U, \quad (12)$$

$$\sum_{g \in M} \sum_{l \in SC} x_{glk} \geq 1, \forall k \in V, M \subset W, \quad (13)$$

$$\sum_{p \in W} g_{pr} \geq u_r, \forall r \in S, \quad (14)$$

$$g_{pr} \leq u_r, \forall p \in W, r \in S, \quad (15)$$

$$\varphi_{irh} \leq u_r, \forall i \in O, r \in S, h \in U, \quad (16)$$

$$\sum_{r \in S} \varphi_{irh} \leq 1, \forall h \in U, i \in O, \quad (17)$$

$$\sum_{i \in W} \sum_{j \in SC} x_{ijk} \leq 1, \forall k \in V, \quad (18)$$

$$\sum_{k \in V} \sum_{j \in SC} x_{pj k} \geq z_p, \forall p \in W, \quad (19)$$

$$\sum_{j \in SC} x_{pj k} \leq z_p, \forall k \in V, p \in W, \quad (20)$$

$$\sum_{g \in W \cup SC} x_{jgk} + \sum_{g \in W \cup SC} x_{pgk} - y_{jp} \leq 1, \forall j \in SC, k \in V, p \in W, \quad (21)$$

$$ST_{jk} = ST_{ik} + \tau_{ijk} x_{ijk}, \forall i, j \in W \cup SC, k \in V, \quad (22)$$

$$BT_j = DT \times [q_j - \text{rep}(q_j, QV)] \times d_{pj} \times y_{jp}, \forall j \in BC, p \in W, \quad (23)$$

$$ET_{jh} = ET_{ih} + \zeta_{ijh} \varphi_{ijh}, \forall i, j \in O \cup S \cup SC, h \in U, \quad (24)$$

$$d_{jh} = m(1 - \alpha_{ij}) + \tilde{\delta}_{ij}, \quad (25)$$

$$\sum_{j=1}^2 w_j = 1, \quad (26)$$

$$x_{ijk} \in \{0, 1\}, \varphi_{ijh} \in \{0, 1\}, u_r \in \{0, 1\}, z_p \in \{0, 1\}, g_{pr} \in \{0, 1\}, y_{jp} \in \{0, 1\}, f_{s_p} \geq 0, f w_{pj} \geq 0, \forall i, j \in C, k \in V, h \in U, p \in W, r \in S. \quad (27)$$

Among them, the objective function shows as follow: Equation (1) indicates the minimal sum of the time spent on distributing relief materials to each settlement (shelters) and road repairing. Equation (2) indicates the minimal total cost of post-earthquake emergency logistics

system, including the cost of facilities establishment, means of dispatching transportation (vehicles and helicopters), relief materials distribution and road repairing. Constraint condition: Equation (3) indicates the volume of materials sending out from the collector-distributor point is less than the capacity of this collector-distributor point. Equation (4) indicates the volume of materials sending from collector-distributor point to each emergency distribution centres is less than the capacity of this emergency distribution centre. Equation (5) indicates transport volume conservation. Equation (6) indicates that the possibility of total demand of all the settlements (shelters) assign to the emergency distribution centre p does not exceed its capacity should maintain in the confidence level. Equation (7) indicates that the possibility of the total demand of small settlements (shelters) assign to the circuted (touring) vehicle k does not exceed the capacity of this vehicle should maintain in the confidence level. Equation (8) indicates that the relief materials distributed to the big settlement (shelter) j ($j \in BC$) should meet the demand of this settlement (shelter), $\text{rep}(q_j, VQ)$ indicates the remainder of large settlement demand and the capacity of maximum cargo vehicle. Equation (9) indicates that the emergency helicopter distribution can only be used at the last settlement (shelter) with unmet needs of a route. Equation (10) indicates that the unmet demand of each settlement (shelter) assign to helicopter h cannot exceed the capacity of this helicopter. Equation (11) and equation (12) indicate the continuous constraints of route. Equation (13) indicates constraints of eliminating circuted distribution, which means each routes should connect to at least one emergency distribution centre. Equation (14) and equation (15) indicate that once the collector-distributor point is established, the emergency distribution centre will assign to it, and the emergency distribution centre only assign to those open collector-distributor point. Equation (16) and equation (17) indicates that the helicopter can only be set out to the open collector-distributor point from the airport, and each helicopter can only go to one collector-distributor point. Equation (18) indicates that each vehicle can only assign to one emergency distribution centre at the most. Equation (19) and equation (20) indicates that once the emergency distribution centre is established, there will be vehicles assign to it, and the vehicles can only be assigned to the open emergency distribution centre. Equation (21) indicates that the small settlement (shelter) j ($j \in SC$) can assign to the emergency distribution centre p only when there is one and only route pass this small settlement (shelter) from centre p . Equation (22) to Equation (24) show the mathematical expressions of ST_{jk} , BT_j and ET_j . Equation (25) indicates the Euclidean distance transforms into actual distance. Equation (26) indicates that the sum of the weight of all targets is 1. Equation (27) is the constraint of 0-1 decision variable and the constraint of nonnegative number.

Constraint equation (6) and equation (7) cannot be figured out directly because there is fuzzy parameter q_j in it. They require for proper processing. Assume that ΔVQ_θ is the remaining capacity of rescue vehicle k after it finish the service of several (the number is θ) settlements (shelters). ΔWQ_ε is the remaining capacity of emergency distribution centre p when it finishes the service of several (the number is ε) settlements (shelters). ΔVQ_θ and ΔWQ_ε are Triangular fuzzy numbers (Jian-yong Zhang [13]), there are:

$$\Delta VQ_\theta = (VQ_k - \sum_{j=1}^{\theta} c_j, VQ_k - \sum_{j=1}^{\theta} b_j, VQ_k - \sum_{j=1}^{\theta} a_j), \quad (28)$$

$$\Delta WQ_\varepsilon = (WQ_p - \sum_{j=1}^{\varepsilon} c_j, WQ_p - \sum_{j=1}^{\varepsilon} b_j, WQ_p - \sum_{j=1}^{\varepsilon} a_j). \quad (29)$$

The possibility that demand of next settlement (shelter) is less than the remaining capacity of the vehicle and demand of next settlement (shelter) is less than the remaining capacity of the emergency distribution centre can be indicated by equation (30) and equation (31) as follow:

$$p_1 = \text{Pos}\{q_{\theta+1} \leq \Delta VQ_\theta\} = \begin{cases} 1, & b_{\theta+1} \leq \Delta VQ_\theta \\ \frac{\Delta VQ_\theta - a_{\theta+1}}{b_{\theta+1} - a_{\theta+1}}, & a_{\theta+1} < \Delta VQ_\theta < b_{\theta+1} \\ 0, & a_{\theta+1} \geq \Delta VQ_\theta \end{cases}, \quad (30)$$

$$p_2 = \text{Pos}\{q_{\varepsilon+1} \leq \Delta WQ_\varepsilon\} = \begin{cases} 1, & b_{\varepsilon+1} \leq \Delta WQ_\varepsilon \\ \frac{\Delta WQ_\varepsilon - a_{\varepsilon+1}}{b_{\varepsilon+1} - a_{\varepsilon+1}}, & a_{\varepsilon+1} < \Delta WQ_\varepsilon < b_{\varepsilon+1} \\ 0, & a_{\varepsilon+1} \geq \Delta WQ_\varepsilon \end{cases}. \quad (31)$$

The vehicle will go on visiting next node when $p_1 \geq \tilde{p}$, otherwise it will return to the emergency distribution centre, as $x_{ijk} = 1 \Rightarrow \text{Pos}\{q_j \leq \Delta VQ_i\} \geq \tilde{p}$. This emergency distribution centre can accept next settlement (shelter) when $p_2 \geq \bar{p}$, as $y_{jp} = 1 \Rightarrow \text{Pos}\{q_j \leq \Delta WQ_{j-1}\} \geq \bar{p}$. Therefore, constraint equation (6) and equation (7) are separately equivalent to $x_{ijk} = 1 \Rightarrow \text{Pos}\{q_j \leq \Delta VQ_i\} \geq \tilde{p}$ and $y_{jp} = 1 \Rightarrow \text{Pos}\{q_j \leq \Delta WQ_{j-1}\} \geq \bar{p}$.

3 Algorithm design

Aforementioned model is a fuzzy multi-objective mixed integer programming model, here we proposed an improved genetic algorithm based on weight coefficient transform, that is transform the multi-objective optimization problem into a single objective optimization

problem. The solution of the LRP algorithms were formerly based on two stage solution ideas or using the intelligence algorithm [11]. However, this thesis designs a genetic algorithm to solve the LRP as a whole. This can effectively reduce the probability of stagnating in the local optimal solution in the process of evolution. Besides, in order to prevent the premature problem of genetic algorithm, this algorithm adopts stochastic universal sampling, restructuring strategy, variable cross rate and variation rate method.

3.1 THE ALGORITHM PARAMETER SETTING

Assume that Pop is the population size, GAPP is the generation gap probability, Maxgen is the Maximum repeatedly algebra, P_c is the probability of crossover; P_m is the probability of metamorphosis.

3.2 CODING AND INITIAN

It needs to use specific encoding method. Each chromosome consists of six substrings; there are n gene positions in the first substring. n stands for the number of settlements(shelters). This substring corresponds to numbered settlements (shelters); the value of each gene position is a random number in from 1 to K . K stands for the number of vehicles. As for the big settlements (shelters), here we only consider remaining demand after the full-load distribution. In addition, there are n gene positions in the second substring, which correspond to numbered settlements (shelters). The difference is that this part stands for the full-load capacity from distribution centre to the big settlements (shelters) (that is the demand of big settlements minus the remainder of its demand divided by the capacity of maximum vehicle). The value of each gene position is a natural number in from 1 to P , P stands for the number of alternative emergency distribution centres. Length of the third substring is K , the gene positions of this substring correspond to numbered vehicles, the value of each gene position is a natural number in from 1 to P . Length of the fourth substring is $R \times P$, there are $R \times P$ gene positions, separately correspond to the numbered emergency distribution centres, $((r-1) \times p + 1$ to $r \times p$ indicates the distribution centres assigned to the collector-distributor point r , thus each distribution centre assign not only to one collector-distributor point, but the number must be less than the number of established collector-distributor points). The value of each gene position is a natural number in from 1 to R , R stands for the number of alternative relief materials collector-distributor points. Length of the fifth substring is n , this substring corresponds to the numbered demand-unmet settlements (shelters). In addition, the value of each gene position is a random number in from 1 to H , H stands for the number of the helicopter. Length of the sixth substring is H , the value of each gene position is a natural number in from 1 to R . Using this kind of encoding method can implicate the information of

guaranteeing every settlements (shelters) to get service, vehicles distributed to the emergency distribution centres, distribution centres be assigned to collector-distributor points and which collector-distributor point will a helicopter take off from during emergency distribution, this will be convenient for dealing with every constraint in the model.

3.2.1 Constraint of the capacity

The penalty function method is applied for dealing with the constraint of the capacity of conveyance, emergency distribution centres and emergency materials collector-distributor points. That is to multiply a penalty coefficient to the objective function of the individual, which does not satisfy the constraint.

3.2.2 Constraint of the settlements (shelters)' demands

Divide the demand of high-demand settlements (shelters), distribute several vehicles to make full-load distribution and use the circuted (tour) vehicles to satisfy the remaining demand.

3.2.3 Other constraints

Use specific code so that the search space and the solution space can match each other. Besides, using genetic algorithm breeding device to mutate the operator can guarantee that the new individual which comes from the mutation maintaining in the corresponding solutions in the solution space.

The genetic algorithm universal function is developed by the University of Sheffield in Britain to set up the initial population. First of all, use *crtbase* function to establish 3 vector quantity: The length of a basic character vector quantity is $2n$, it is combined with n basic characters $\{0,1,2,\dots,K-1\}$ whose cardinal number is K and n basic characters $\{0,1,2,\dots,P-1\}$ whose cardinal numbers is P . (n stands for the number of the basic characters). Here is the second basic character vector quantity; its length is $K+R\times P$. It's combined with K basic characters $\{0,1,2,\dots,P-1\}$ whose cardinal number is P and $R\times P$ basic characters $\{0,1,2,\dots,R-1\}$ whose cardinal number is R . (K and $R\times P$ stand for the number of the basic characters). The third basic character vector quantity with a length of $n+H$, it is combined with n basic characters $\{0,1,2,\dots,H-1\}$ whose cardinal number is H and H basic characters $\{0,1,2,\dots,R-1\}$ whose cardinal number is R . (n and H stand for the number of the basic characters.) Secondly, using *crtbp* function to establish 3 matrices *Chrom1*, *Chrom2* and *Chrom3* with the same line number and their elements are random number (their basic characters are determined by the corresponding vector quantity). The line number of the matrices is the population size-Pop. Then we define $Chrom = [Chrom1, Chrom2, Chrom3]$. In order to ensure that the number of the column and the line is not zero in the matrices, we

add a 1-matrix with the same dimension and all its entry is 1. Here we get the initial population.

3.3 CALCULATING THE DEGREE OF ADAPTATION

In the calculation of target value, due to the different targets with different dimension and magnitude, we cannot do direct weighting. In order to eliminate the influence of different dimension and magnitude, it is necessary to do dimensionless processing. Method of dimensionless processing are many, commonly used methods are standardization processing method, extreme value method, the linear scaling method, normalized method, vector norm method and effect coefficient method [14]. This thesis uses linear scaling method, the target value after dimensionless processing is $f'_j = f_j / f_j^0$, $j=1,2$, f_j is the actual value of target j , f_j^0 is the satisfaction value of target j .

Assume that $u_i = \text{Fitness}(Chrom_i(gen))$ is the fitness degree of the No. i chromosome of the gen generation population-Chrom. It is figured out by the transform of the dimensionless target function. So the fitness degree function is $u_i = 1 / (w_1 f'_1(i) + w_2 f'_2(i))$. $f'_j(i)$ is the evaluation of the target j of chromosomes i after dimensionless processing. w_j is the weight of no. j target, it can be given different weight according to the degree of importance of the objective function. In the post earthquake emergency response phase, time is life, so the weight in the target weight is larger. Cost is also a factor which needs to be considered when the emergency resources are very limited but the urgent demand for them is very large. Compare with the weight of time, weight of cost is smaller.

3.4 GENETIC OPERATION

3.4.1 Select operation

This study uses the stochastic universal sampling and fitness reinsertion method. Stochastic universal sampling can make the population diversity relatively permanent, and prevent the algorithm premature convergence. However, fitness reinsertion will make the most appropriate individuals to be generational reproduced continuously. This can improve the effectiveness of the optimization strategy.

3.4.2 Crossover operation

In order to maintain the diversity of population, this study adopt the restructuring strategy based on crossover operator, adopt the crossover operator which has great destructive – multi point crossover operator. This can promote the search of the solution space by the genetic

algorithm. Besides, variable cross rate Pc needs to be used.

3.4.3 Mutation operation

Using the real variation, and add in the region descriptor to limit the range of variation, in order to ensure that the mutation does not exceed the decision variable boundary. Variable mutation probability in the selection of P_m in this study represents the probability of mutation. That is to use larger value for P_m in the earlier stage of algorithm to expand the search space, and use smaller value for P_m in the later stage of algorithm to speed up convergence.

3.5 THE TERMINATE OF ALGORITHM

If $gen \leq Maxgen$, repeating step(3)~step(4). If $gen > Maxgen$, output the optimal chromosome and the optimal solution, the algorithm terminates.

4 Example analysis

Randomly presents 2 alternative relief materials distribution points, 4 alternative emergency distribution

centre, 10 vehicles, 3 helicopters and airport location (coordinates (50,285)), 30 settlements (shelters). Getting the fuzzy demand and the coordinates of each settlements (shelters) according to some available information (as shown in the table 1). The plane coordinates of the emergency facilities and the settlements (shelters) are two-dimension random numbers from 0 to 300. This study uses the simulating method from Zhang et al to make "actual" demand, which comprises the following steps: (1) Generating a random number x in the fuzzy demand number of certain settlement (shelter), and calculate the degree of its relative-membership μ_{q_j} . (2) Generating a random number ϑ in the range of $[0, 1]$. (3) Comparing ϑ and μ_{q_j} , if δ is smaller than μ_{q_j} , x will be the actual demand of this settlement(shelter), otherwise, repeat the above steps, until we get the actual demand of this settlement (shelter). (4) Repeating the above steps and generate the actual demand of all the settlements (shelters).

TABLE 1 Coordinate of resettlement sites and its fuzzy demand of relief materials

Serial Number	1	2	3	4	5	6	7	8	9	10
Coordinate	(10,51)	(21,32)	(102,205)	(74,86)	(61,25)	(85, 142)	(56,164)	(112,55)	(93,126)	(124,175)
Fuzzy Demand	22,25,30	32,36,40	35,38,42	38,40,45	42,45,48	34,36,40	52,55,60	30,36,40	42,46,48	50,52,55
Serial Number	11	12	13	14	15	16	17	18	19	20
Coordinate	(46,105)	(185,88)	(92,19)	(205,184)	(225,115)	(282,156)	(253,224)	(154,245)	(248,38)	(300,215)
Fuzzy Demand	40,43,45	52,55,58	52,56,60	640,644,650	60,62,65	82,85,88	42,45,50	62,64,68	52,56,60	82,85,88
Serial Number	21	22	23	24	25	26	27	28	29	30
Coordinate	(150,27)	(160,55)	(210,25)	(240,50)	(40,210)	(50,150)	(60,250)	(125,80)	(100,75)	(250,150)
Fuzzy Demand	42,44,48	62,65,63	54,56,60	79,82,85	63,65,68	28,30,35	32,35,38	42,45,48	62,64,68	42,46,50

Assume the conversion coefficient $\delta = 1.5$, randomly give α_{ij} the value of 0 or 1, and $\alpha_{ij} = \alpha_{ji} (i, j \in N)$. β_{ij} is a random non negative real number which is between 0 to 1. $\eta = 0.05$ yuan/km, $\gamma = 2$ h/km, $CB=2$ yuan/item·km, $D=1$ yuan/km, $A=5$ yuan/km, $DT=0.001$ h/item·km, \bar{p} and \tilde{p} are all equal to 0.9. Capacity of each vehicle (item), dispatch money (yuan/vehicle), speed(km/h) separately come from set {200, 200, 180, 200, 200, 200, 200, 160, 200, 200}, {2000, 2000, 1800, 2000, 2000, 2000, 2000, 1600, 2000, 2000} and {80, 80, 70, 80, 70, 80, 80, 60, 80, 80}. Capacity of helicopter, dispatch money and speed come from set {50, 50, 50}, {10000, 10000, 10000} and {200, 200, 200}. The coordinates, capacity and the cost of construction of each collector-distributor point and emergency distribution centre come from set {(20, 268), (255, 290)}, {2800, 2600}, {28000, 26000} and set {(55, 52), (100,175), (255,165), (152,125)}, {1000, 1000, 800, 800}, {10000, 10000, 8000, 8000}. The unmet demand of the last visited settlements (shelters) of each vehicle route come from {0, 8, 4, 2, 7, 1, 0, 1, 1, 2, 3, 9, 8, 7, 6, 10, 5, 8, 5, 5, 2, 5, 7, 4, 8, 6, 7, 0, 4, 2}.

Assume Pop (population size) = 100, GAPP (generation gap probability) = 0.9, maxgen (Maximum repeatedly algebra) = 200, p_c (probability of crossover) in the earlier stage is 0.5, in the later stage is 0.7. p_m (probability of metamorphosis) in the earlier stage is 0.09, in the later stage is 0.02. Weight of time and cost are respectively 0.8 and 0.2. According to the above algorithm, using MATLAB M language programming calculation, running the program 10 times in the Intel CoreTM 2 Duo 1.83G 1G memory notebook computer, and then selecting the best results. The weighted objective value is 18.24, program running time is 181.86s, the LRP optimization scheme as shown in Figure 1. Figure 2 shows the situation of the population optimal solution of genetic algorithm and the variation of the mean value.

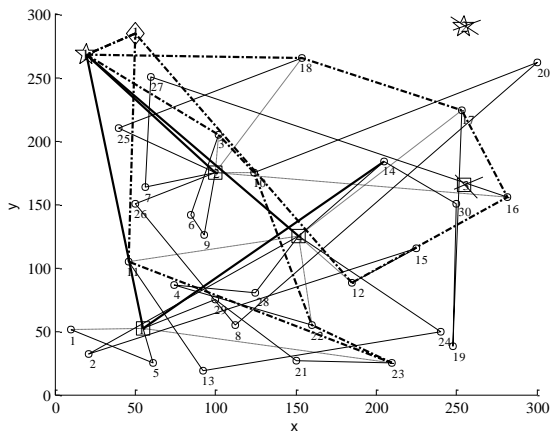


FIGURE 1 One Solution for LRP

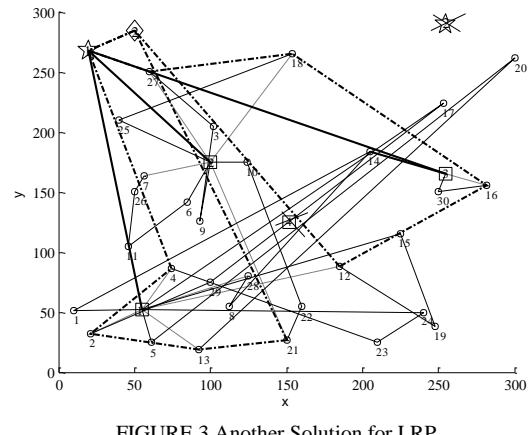


FIGURE 3 Another Solution for LRP

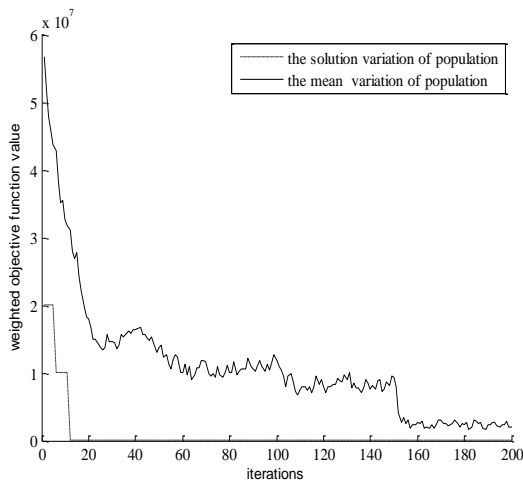


FIGURE 2 A Capability Track Chart of Algorithm

Furthermore, this study tests the robustness of algorithm. We record the optimal value of 10 times operations when the algorithm parameters are fixed. Those operations include 18.57, 19.29, 19.49, 21.05, 19.76, 18.24, 20.71, 18.48, 18.99, and 22.49. The average value of them is 19.70, and the absolute value of the difference between those 10 times operations' optimal value and the average value are respectively 1.14, 0.42, 0.22, 1.34, 0.05, 1.45, 1.00, 1.22, 0.72, 2.78. These values don't have large difference, so the design of the improved genetic algorithm has strong robustness.

When P_c and P_m is fixed, extended population size is 200, maxgen is 400, running for 5 times to get the best result 14.94, the LRP optimization scheme as shown in figure 3. This indicates that the population size and the maximum repeatedly algebra were increased to 200, 400, its target value only decreased a little. Therefore, satisfactory result has appeared in the 100 population size and 200 generation, to increase those two parameters (the population size and the maximum repeatedly algebra) will not make the objective function obviously improved.

5 Conclusions

This study establishes an optimized multi objective fuzzy LRP model, which aims at minimizing the total time of emergency logistics system operation and the total cost of emergency logistics system. We also propose an improved genetic algorithm based on weighted coefficient transformation. After that, we adopt a special real valued coding scheme, the penalty function method and the materials demand segmentation strategy to handle the constraints in the model. The research adopt stochastic universal sampling, restructuring strategy, variable cross rate and variation rate method, in order to prevent the premature convergence of genetic algorithm. Example analysis shows that this model and algorithm can effectively solve the fuzzy LRP of the post-earthquake emergency logistics system.

This study considers about the fuzzy demand of the relief materials of the settlements (shelters), the final (vehicle) visited settlement(shelter) of each route, the possibility of unmet demand, the need of emergency dispatching helicopters from the collector-distributor points to fulfil the needs of this settlement (shelter), and then constructs a single variety emergency materials distribution LRP model. However, as the variety of post-earthquake emergency materials is diverse, the multi-variety LRP needs further research. Besides, the uncertainty of the rescue vehicle's travel time, the time window of the demand of emergency materials and other factors should also be considered.

Acknowledgments

The research was supported by National Natural Science Foundation of China (No.90924012, No.71090402, No.71103149); Program for New Century Excellent Talents in University (No.NCET-10-0706); the Ministry of education of Humanities and Social Science Research Fund Project (08JC630067; 12YJC630210); Sichuan Province Youth Science and Technology Foundation (09ZQ026-021); Sichuan provincial leaders in academic and technical training fund project (Sichuan people

commune No. [2011]441); Fundamental Research Funds for the Central Universities (No.SWJTU11CX152, No.2682013CX073); Huaqiao University research start-up cost project (11BS118).

References

- [1] Min H, Jayaraman V, Srivastava R 1998 Combined location-routing problems: A synthesis and future research direction *European Journal of Operational Research* **108**(1) 1-15
- [2] Nagy G, Salhi S 2007 Location-routing: Issues, models and methods *European Journal of Operational Research* **177**(2) 649-72
- [3] Sérgio B, Carlos F, José P 2007 Using clustering analysis in a capacitated location-routing problem *European Journal of Operational Research* **179**(3) 968-77
- [4] Haghani A, Oh S 1996 Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations *Transportation Research Part A* **30**(3) 231-50
- [5] Barbarosoğlu G, Arda Y 2004 A two-stage stochastic programming framework for transportation planning in disaster response *Journal of the Operational Research Society* **55**(1) 43-53
- [6] Hang W, Jie W 2009 Emergency path problem in stochastic and time-varying network *Journal of Systems Engineering* **24**(1) 99-103 (in Chinese)
- [7] Yuan Y, Dingwei W 2009 Path selection model and algorithm for emergency logistics management *Computer & Industrial Engineering* **56**(3) 1081-94
- [8] Sheu J 2007 Challenges of emergency logistics management *Transportation Research Part E* **43**(6) 655-9
- [9] Sheu J 2010 Dynamic relief-demand management for emergency logistics operations under large-scale disasters *Transportation Research Part E* **46**(1) 1-17
- [10] Yi W, Özdamar L 2007 A dynamic logistics coordination model for evacuation and support in disaster response activities *European Journal of Operational Research* **179**(3) 1177-93
- [11] Shaoren W, Zujun M 2011 Location-routing problem in emergency logistics system for post-earthquake emergency relief response *Systems Engineering-Theory & Practice* **31**(8) 1497-507 (in Chinese)
- [12] Tomasini R M, Wassenhove L N 2004 Pan-American Health Organization's Humanitarian Supply Management System: Depoliticization of the Humanitarian Supply Chain by Creating Accountability *Journal of Public Procurement* **4**(3) 437-49
- [13] Jian-yong Z, Jun L, Yao-huang G 2004 Research of the Fuzzy Dynamic Vehicle Scheduling Problem When Demand at Nodes is Uncertain *Journal of Management Engineering* **18**(4) 69-72 (in Chinese)
- [14] Yajun G, Feng-mei Ma, Qingxing D 2011 Analysis of influence of dimension less-methods on deviation maximization method *Journal of Management Sciences in China* **14**(5) 21-9 (in Chinese)

Authors

	<p>Shaoren Wang, born in January, 1981, Wenshan City, Yunnan Province, P.R. China</p> <p>Current position, grades: the lecturer of College of Business Administration, HuaQiao University</p> <p>University studies: Graduated from College of Transportation and Logistics of Southwest Jiaotong University in 2010, received a doctor's degree.</p> <p>Scientific interest: logistics and supply chain management, emergency management.</p> <p>Publications: Presided over 2 scientific research projects the completion of provincial; more than 10 papers published in various journals.</p>
	<p>Zujun Ma, born in 1974, Kaihua City, Zhejiang Province, P.R. China</p> <p>Current position, grades: professor, doctoral supervisor of School of Economics and Management of Southwest Jiaotong University.</p> <p>University studies: Graduated from School of Economics and Management of Southwest Jiaotong University in 2002, received a doctor's degree.</p> <p>Scientific interest: logistics and supply chain management, emergency management.</p> <p>Publications: Presided over the completion of national and provincial, than 10 scientific research projects; more than 140 papers published in various journals.</p>
	<p>Bochao Zhuang, born in September, 1982, Quanzhou City, Fujian Province, P.R. China</p> <p>Current position, grades: Assistant Professor from Faculty of Business Administration, Huaqiao University.</p> <p>University studies: Graduated from School of Business Administration, The Chinese University of Hongkong in 2012, received a doctor's degree.</p> <p>Scientific interest: logistics outsourcing, supply chain management, quality management.</p> <p>Experience: Graduated from South China University of Technology in Guangzhou in 2007 and received Master degree.</p>

Evaluation of regional economic development level based on grey clustering and rough set

Peizhe Li^{1, 2*}, Lirong Jian¹, Kun Zhang^{1, 2}, Shanshan Pei²

¹ College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, 210016

² School of Business, Shan Dong University of Political science and Law, Jinan, Shandong, China, 250014

Received 8 October 2013, www.tsi.lv

Abstract

The evaluation result of regional economic development is an important basis for government to make regional economic development planning, scientific regional economic development evaluation system and method has an important significance for understanding the regional economic development. Using grey clustering to classify different regions by the new evaluation index system, and using the rough set theory to derive fuzzy decision rules, these rules can explain the preference behaviour of decision makers and provide scientific and rational decision-making suggestions.

Keywords: Regional economic, Grey fixed weight clustering, Dominance rough set, Evaluation, Decision rules

1 Introduction

Regional economy is the foundation of the national economy. It can reflect the development level of economy accurately and systematically as well as objectively measure the economic development process; it has strategic significance to provide accurate and scientific statistical data for the leaders in time. Therefore, it has important practical significance to establish an accurate and systematical index system, which can measure and evaluate the level of regional economic development.

Many scholars have already made an investigation on the evaluation of regional economic development. Song Malin has adopted factor analysis and regression analysis to evaluate the economic development of Anhui province comprehensively [1]; Luan Jinchang etc. has adopted the entropy theory to determine the index weight, then used AHP to evaluate the cities level economic [2]; Wang Xiaoliang has constructed the evaluation index system including 24 indicators of regional open economy development level from 4 aspects: opening up foundation, scale, architecture, and efficiency, taking the seven provinces and municipalities in the eastern coastal with higher degree of opening as an example, adopting the improved entropy method to evaluate and sort the level of the seven provinces of the open economy in 2010 [3]; Sheng Minglan has constructed the evaluation index system and evaluation model for the county (district) level of economic development. Based on AHP and grey correlation method, She has got the multi-level grey evaluation model and made use of three aspects of economic strength, prosperity, development speed so as

to evaluate the domain of the economic development level of the 40 counties (District) in Chongqing city[4]; Wang Xizhao etc., based on the five aspects of scale as economic stock index, benefit index, structure index, pressure index and prosperity index, has adopted factor analysis method to evaluate the level of economic development of Qianxi County, and clustered using SAS program [5].

Above all, there are many researches on the existing literature concern regional economic development level, but many evaluation methods are quite subjective and relatively simple, which subjectively determine the weights of the evaluation without simplify indicators. Therefore, an objective evaluation of the regional economic development and the deriving of decision rules will create great practical significance. This paper has adopted entropy method to determine the weight of indicators, adopted the method of grey fixed weight clustering to classify the regional economic development level of 17 cities of Shandong, and derived the decision rules by using the method of rough set. It can provide the important reference basis to the regional economic development for the local government to make the planning and decision.

2 Grey clustering and rough set theory

In the face of uncertainty decision problems, if policymakers want to make quick decisions, they have to prepare effective analysis and processing tools while rough set theory is a powerful tool for data processing. Rough set theory can deal with imprecise, uncertain and

*Corresponding author - Tel: +86-139-5319-2873 fax: +86-139-5319-2873; E-mail: lipeizhe713@163.com

incomplete data. It can effectively analyse incomplete information, which has been successfully applied to knowledge acquisition, decision analysis, prediction, expert system and knowledge discovery in database and many other fields. However, with some of disadvantages, the rough set theory cannot solve every problem. Only when it has combined some other theories such as grey system, neural network, probability statistics, fuzzy sets of two or more complementary heterozygous and build more powerful hybrid soft decision method, expand the rough set theory of applicable scope, can it deal with uncertainty decision problems effectively. However, the decision table of rough set must be given in advance, and we can only establish information system based on dataset without acquire the decision table. Besides, clear approximation space will always be met, and the decision attribute values are always fuzzy. As a result, we could not obtain probabilistic decision rules [6-8]. This paper has adopted a hybrid method of rough sets and grey clustering. Generate the decision table of rough fuzzy sets by using grey clustering method, then reduction and approximate method for the attribute so as to acquire decision rules.

2.1 GREY FIXED WEIGHT CLUSTERING

Grey clustering is method based on grey correlation matrix or grey whitenization weight function so as to divide some observation index or observed object into defined categories.

Let $x_{ij}(i=1,2,\dots,n; j=1,2,\dots,m)$ as the observation object i index of the j value, $f_j^k(\cdot)(j=1,2,\dots,m; k=1,2,\dots,s)$ as k whitenization weight function of j index. If the weight $\eta_j^k(j=1,2,\dots,m; k=1,2,\dots,s)$ of subclass k are independent with k , for arbitrary $k_1, k_2 \in \{1,2,\dots,s\}$, always has $\eta_j^{k_1} = \eta_j^{k_2}$, the k of η_j^k can be omitted, denoted as $\eta_j(j=1,2,\dots,m)$, and $\sigma_i^k = \sum_{j=1}^m f_j^k(x_{ij})\eta_j$ is called the grey fixed weight clustering coefficient that the object i belongs to the grey type k ; if for any $j=1,2,\dots,m$, always has $\eta_j = \frac{1}{m}$, then called

$\sigma_i^k = \frac{1}{m} \sum_{j=1}^m f_j^k(x_{ij})$ grey equal weight clustering coefficient for object i belongs to the k grey class.

Clustering objects are classified according to the grey fixed weight clustering coefficient called grey fixed weight clustering.

The steps of Grey fixed weight clustering:

(1) Gives the subclass of k of j indicators' whitenization weight function

$f_j^k(\cdot)(j=1,2,\dots,m; k=1,2,\dots,s)$.

(2) Determine the index clustering weight $\eta_j(j=1,2,\dots,m)$.

(3) From the 1, 2 step of the whitenization weight function of $f_j^k(\cdot)(j=1,2,\dots,m; k=1,2,\dots,s)$, the clustering weight $\eta_j(j=1,2,\dots,m)$ and observations $x_{ij}(i=1,2,\dots,n; j=1,2,\dots,m)$ of the object i index of j , calculated the grey fixed weight clustering coefficient

$$\sigma_i^k = \sum_{j=1}^m f_j^k(x_{ij})\eta_j \quad (j=1,2,\dots,m; k=1,2,\dots,s).$$

(4) Get the clustering coefficient vector $\sigma_i = (\sigma_i^1, \sigma_i^2, \dots, \sigma_i^s)$ by grey fixed weight clustering coefficient.

(5) Clustering coefficient matrix

$$\Sigma = (\sigma_i^k)_{n \times s} = \begin{bmatrix} \sigma_1^1 & \sigma_1^2 & \dots & \sigma_1^s \\ \sigma_2^1 & \sigma_2^2 & \dots & \sigma_2^s \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^1 & \sigma_n^2 & \dots & \sigma_n^s \end{bmatrix}. \quad (1)$$

(6) If $\max_{1 \leq k \leq s} \{\sigma_i^k\} = \sigma_i^{k^*}$, then determine the object i belongs to the grey k^* .

2.2 DOMINANCE ROUGH SET THEORY

In many problems, learning classification is a class of problems that has been widely studied. Rough set theory is a powerful tool for reasoning about data. And learning system, based on rough set theory, is the essence of that classification problem given attributes redundancy and dependence. Rough set theory is based on such assumption; each object and some information in the domain of discourse were linked while the same amount of information describing the object cannot be distinguished. The indiscernible relation generated by this manner is a mathematical basis of rough set theory. However, in many actual problems, the consideration of sequence characteristics of preference is important [9-10]. With a dominance relation instead of indiscernible relation, an important consequence of this treatment is that it can derived preference model by case decision.

2.2.1 Dominance relation

Let $x, y \in U$, $P \subseteq C$, if for $\forall q \in P, f(y, q) \geq f(x, q)$, then the relationship of $yD_p x$ is called the dominance relation. This definition of dominance relation actual is weak dominance or partial order relation, expression of the object set in preference a preference on q . Give $P \subseteq C$ and $x \in U, y \in U$. On the P -dominating set and P -dominated set of x respectively:

$$D_p^+(x) = \{yD_p x\}, \tag{2}$$

$$D_p^-(x) = \{xD_p y\}. \tag{3}$$

Assuming decision-makers recommend that, the object set U is divided into a finite number of decision classes, let $Cl = \{Cl, t \in \{1, 2, \dots, n\}\}$, then $Cl_n \succ \dots \succ Cl_t \succ \dots \succ Cl_1$, class to set union and downward respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, t, s \in \{1, 2, \dots, n\}. \tag{4}$$

The statement $x \in Cl_t^{\geq}$ means " x at least belongs to the category of Cl_t "; the statement $x \in Cl_t^{\leq}$ means " x most belongs to the category of Cl_t ". From here we can see, the each object of the up union Cl_t^{\geq} is better than or at least equal to each object of the down union Cl_t^{\leq} . Obviously, $Cl_1^{\geq} = Cl_n^{\leq} = U, Cl_n^{\geq} = Cl_1^{\leq}$. In addition, for $t = 2, \dots, n$, then $Cl_{t-1}^{\leq} = U - Cl_t^{\geq}, Cl_t^{\geq} = U - Cl_{t-1}^{\leq}$. In other words, policymakers have made decision table object assigned to Cl categories according to the following of the comprehensive evaluation: the difference in the Cl_t class, the best in class Cl_n , and other objects belonging to the rest of the class Cl_t . According to this principle, if $t \in \{1, 2, \dots, n\}$ is larger, then the category Cl_t is better.

2.2.2 Dominate rough set approximation

Assuming the knowledge represents system $S = (U, A, V, f)$ as partial multi-attribute decision table, $A = C \cup D$, given the preference information's attribute set $P \subseteq C, x \in U, Cl_t \subseteq D, t \in \{1, 2, \dots, n\}$. According to Greco's theory, the lower and upper approximation of Cl_t^{\geq} respectively:

$$\underline{apr}_p(Cl_t^{\geq}) = \bigcup \{x \in U : D_p^+(x) \subseteq Cl_t^{\geq}\}, \tag{5}$$

$$\overline{apr}_p(Cl_t^{\geq}) = \bigcup \{x \in U : D_p^-(x) \cap Cl_t^{\geq} \neq \emptyset\}. \tag{6}$$

$\underline{apr}_p(Cl_t^{\geq})$ makes up of all objects of Cl_t^{\geq} set, \overline{apr}_p makes up of all objects of Cl_t^{\geq} set.

The boundary area for Cl_t^{\geq} :

$$bnd_p(Cl_t^{\geq}) = \underline{apr}_p(Cl_t^{\geq}) - \overline{apr}_p(Cl_t^{\geq}). \tag{7}$$

Similarly, the lower and upper approximation of Cl_t^{\leq} respectively:

$$\underline{apr}_p(Cl_t^{\leq}) = \bigcup \{x \in U : D_p^-(x) \subseteq Cl_t^{\leq}\}, \tag{8}$$

$$\overline{apr}_p(Cl_t^{\leq}) = \bigcup \{x \in U : D_p^+(x) \cap Cl_t^{\leq} \neq \emptyset\}. \tag{9}$$

$\underline{apr}_p(Cl_t^{\leq})$ is composed by all objects of Cl_t^{\leq} , \overline{apr}_p is composed by all objects of Cl_t^{\leq} .

The boundary area for Cl_t^{\leq} :

$$bnd_p(Cl_t^{\leq}) = \overline{apr}_p(Cl_t^{\leq}) - \underline{apr}_p(Cl_t^{\leq}). \tag{10}$$

The element boundary region of Cl_t^{\leq} or Cl_t^{\geq} cannot explain its ownership.

2.2.3 Classification and reduction

The classification quality of Cl :

$$r_p(Cl) = \frac{|U - ((\bigcup bnd_p(Cl_t^{\geq})) \cup (\bigcup bnd_p(Cl_t^{\leq})))|}{|U|}. \tag{11}$$

The $r_p(Cl)$ express ratio of a lot of object attributes in decision table of correct classification number and the total number of object.

The minimal subset $P \in C$, which meet the $r_p(Cl) = r_c(Cl)$ called a reduction of C about Cl , expressed as $red_{Cl}(P)$. A preference decision table may have more than one reduction, the intersection of all reduction called core. The core is the most important attribute preference of attributes in decision table set, it may be empty also.

2.2.4 Preference decision rules

Preference decision rule is a form between conditional preference attributes and decision preference attributes. Based on rough approximation of dominance relation, the preference decision rules in decision table will be derived from the preference information. The decision rule set can be divided into decision classes set $D \geq$ set of decision rules and decision class down and $D \leq$ decision rule set, form decision rules are generated for $D \geq$:

$$f(x, q_1) \geq r_{q_1} \wedge f(x, q_2) \geq r_{q_2} \dots \wedge f(x, q_p) \geq r_{q_p} \text{ then } x \in Cl_t^{\geq}. \tag{12}$$

These rules approximate only by the category to the union of Cl_t^{\geq} support.

Form $D \leq$ decision rule:

$$f(x, q_1) \leq r_{q_1} \wedge f(x, q_2) \leq r_{q_2} \dots \wedge f(x, q_p) \leq r_{q_p} \text{ then } x \in CI_t^{\leq} \quad (13)$$

These rules approximate only by the downward and the category CI_t^{\leq} support.

Among them:

$$\{q_1, q_2, \dots, q_p\} \subseteq C, (r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}, t \in \{1, 2, \dots, n\} \quad (14)$$

3 Empirical analyses

This paper established the database according to the relevant data in 2012 including 15 index variables of 17 cities in Shandong Province. The data is from the China Statistical Yearbook, China Regional Economic City Economic database. According to the method and steps of processing, it has classified economic development level of the 17 cities of Shandong Province, and derived the decision rules.

3.1 DETERMINE THE INDEX WEIGHT

According to the original data, the method of entropy to calculate the weight of each index

TABLE 2 Knowledge representation system of regional economic evaluation

U	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅
n ₁	B	C	B	B	C	C	C	C	B	C	B	B	C	A	A
n ₂	A	C	A	A	A	A	A	B	A	B	A	A	A	A	A
n ₃	C	C	C	C	B	B	C	B	C	B	C	C	D	B	C
n ₄	D	D	D	D	D	D	D	D	D	D	D	D	D	C	D
n ₅	C	A	D	D	D	C	A	B	C	A	C	D	D	A	D
n ₆	B	C	C	B	B	A	D	A	B	A	A	C	C	B	B
n ₇	C	D	C	C	A	B	D	C	B	B	B	C	D	B	B
n ₈	C	D	C	C	B	D	D	C	C	B	C	D	D	B	C
n ₉	C	D	D	D	D	D	D	D	D	D	C	D	D	C	C
n ₁₀	D	B	D	D	D	C	D	D	D	D	C	D	D	C	D
n ₁₁	D	D	D	D	D	D	D	D	D	D	D	D	D	B	D
n ₁₂	D	D	D	D	D	D	D	D	D	D	D	D	D	B	D
n ₁₃	C	D	D	C	B	C	D	D	D	D	C	C	D	C	C
n ₁₄	D	D	D	D	B	D	D	D	D	C	C	D	D	D	D
n ₁₅	D	D	D	D	C	C	D	D	D	C	D	D	D	D	D
n ₁₆	D	D	D	D	D	D	D	D	D	C	D	D	D	C	D
n ₁₇	D	D	D	D	C	D	D	D	D	D	D	D	D	D	D

3.3 GENERATE REGIONAL ECONOMIC DECISION TABLE BY USING GREY CLUSTERING AND ROUGH SET METHOD

In this paper, the evaluation order of dominance of regional economic development level given are $A > B > C > D$, comments $V = (1, 8, 15, 31)$, the indicators are divided into three classes according to the need of practice: The level of regional economic development difference (code 1), the general level of regional economic development (code 2) and the strong level of regional economic development strong class (code 3).

As shown in table 1.

TABLE 1 Regional economic indicator

Index code	Evaluation index	Weight
a1	GDP	0.049
a2	The per capita GDP	0.042
a3	Budgetary revenue of local government	0.071
a4	Budgetary expenditure of local government	0.045
a5	Amount of industrial enterprises	0.059
a6	Gross value of industrial output	0.047
a7	Above scale enterprise product sales taxes and additional	0.159
a8	The total profit above scale enterprises profit	0.063
a9	Total current assets	0.057
a10	Total fixed assets	0.029
a11	Total investment in fixed assets	0.048
a12	Total sales of commodities	0.094
a13	The amount of foreign capital actually used	0.180
a14	Average wages of staff and workers	0.004
a15	The amount of employees at the end of year in the unit	0.052

3.2 ESTABLISH THE KNOWLEDGE REPRESENTATION SYSTEM OF REGIONAL ECONOMIC

According to the index data of different regions, the index will be deal in a discretization way by using SPSS software. Regional economic system can be represented as in table 2.

First, determine the grey on the basis of the specific distribution of decision objectives and indicators.

Grey clustering according to the index data, tectonic whitenization weight function as shown in figure 1.

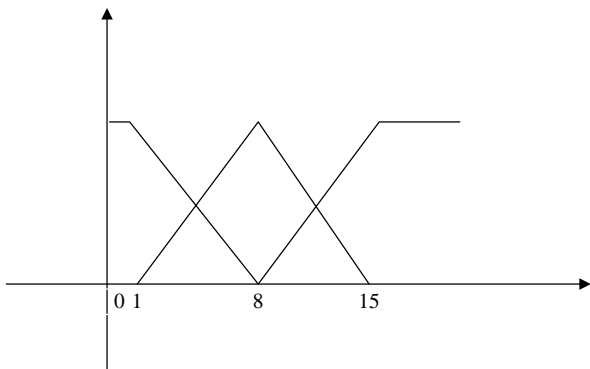


FIGURE 1 Schematic diagram of whitening weight function

Determine the grey according to the whitening weight function:

$$f^1 = \begin{cases} 1, & x \leq 1 \\ -\frac{1}{7}(x-8), & x \in (1,8) \\ 0, & x \geq 8 \end{cases} \quad (15)$$

$$f^2 = \begin{cases} 0, & x \notin [25,75] \\ \frac{1}{25}(x-25), & 25 < x < 50 \\ -\frac{1}{25}(x-75), & 50 \leq x < 75 \end{cases} \quad (16)$$

$$f^3 = \begin{cases} 0, & x < 8 \\ \frac{1}{7}(x-8), & 8 \leq x < 15 \\ 1, & x \geq 15 \end{cases} \quad (17)$$

We can get grey clustering coefficient matrix according to the above formula:

$$R = \begin{pmatrix} 0.0871396 & 0.5514493 & 0.3614111 \\ 0.0006915 & 0.1938554 & 0.8054531 \\ 0.1648041 & 0.7437396 & 0.0914563 \\ 0.73555 & 0.26445 & 0 \\ 0.4129468 & 0.3522875 & 0.2347657 \\ 0.1132877 & 0.5622956 & 0.3244167 \\ 0.0937004 & 0.8131632 & 0.0931364 \\ 0.3421692 & 0.6503078 & 0.0075231 \\ 0.581661 & 0.418339 & 0 \\ 0.5090344 & 0.4814618 & 0.0095039 \\ 0.7747771 & 0.2252229 & 0 \\ 0.9287453 & 0.0712547 & 0 \\ 0.4324709 & 0.5583343 & 0.0091948 \\ 0.6161269 & 0.3838731 & 0 \\ 0.6796908 & 0.3203092 & 0 \\ 0.635886 & 0.364114 & 0 \\ 0.7321114 & 0.2678886 & 0 \end{pmatrix} \quad (18)$$

We can obtain the membership partition table according to the fuzzy clustering coefficient, as shown in table 3.

TABLE 3 Fuzzy membership partition table

U	Fuzzy membership value			The actual decision value
	d_1	d_2	d_3	
n ₁	0.09	0.55	0.36	2
n ₂	0	0.19	0.81	3
n ₃	0.16	0.75	0.09	2
n ₄	0.74	0.26	0	1
n ₅	0.41	0.35	0.24	1
n ₆	0.11	0.56	0.33	2
n ₇	0.09	0.81	0.10	2
n ₈	0.35	0.65	0	2
n ₉	0.58	0.42	0	1
n ₁₀	0.51	0.48	0.01	1
n ₁₁	0.77	0.23	0	1
n ₁₂	0.93	0.07	0	1
n ₁₃	0.43	0.56	0.01	2
n ₁₄	0.62	0.38	0	1
n ₁₅	0.68	0.32	0	1
n ₁₆	0.64	0.36	0	1
n ₁₇	0.73	0.27	0	1

We can obtain regional economic knowledge evaluation decision table according to the clustering coefficient matrix, as shown in table 4.

TABLE 4 Regional economic evaluation decision table

U	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄	a ₁₅	d
n ₁	B	C	B	B	C	C	C	C	B	C	B	B	C	A	A	2
n ₂	A	C	A	A	A	A	A	B	A	B	A	A	A	A	A	3
n ₃	C	C	C	C	B	B	C	B	C	B	C	C	D	B	C	2
n ₄	D	D	D	D	D	D	D	D	D	D	D	D	D	C	D	1
n ₅	C	A	D	D	D	C	A	B	C	A	C	D	D	A	D	1
n ₆	B	C	C	B	B	A	D	A	B	A	A	C	C	B	B	2
n ₇	C	D	C	C	A	B	D	C	B	B	B	C	D	B	B	2
n ₈	C	D	C	C	B	D	D	C	C	B	C	D	D	B	C	2
n ₉	C	D	D	D	D	D	D	D	D	D	C	D	D	C	C	1
n ₁₀	D	B	D	D	D	C	D	D	D	D	C	D	D	C	D	1
n ₁₁	D	D	D	D	D	D	D	D	D	D	D	D	D	B	D	1
n ₁₂	D	D	D	D	D	D	D	D	D	D	D	D	D	B	D	1
n ₁₃	C	D	D	C	B	C	D	D	D	D	C	C	D	C	C	2
n ₁₄	D	D	D	D	B	D	D	D	D	C	C	D	D	D	D	1
n ₁₅	D	D	D	D	C	C	D	D	D	C	D	D	D	D	D	1
n ₁₆	D	D	D	D	D	D	D	D	D	C	D	D	D	C	D	1
n ₁₇	D	D	D	D	C	D	D	D	D	D	D	D	D	D	D	1

3.4 OBTAIN PROBABILISTIC DECISION RULES

We can obtain a reduction by using a rough set attribute reduction algorithm, probabilistic generated by the reduction will determine the rule set, as shown in table 5.

TABLE 5 Probabilistic decision rules

Rule	support	degree of confidence
If GDP <D and Budgetary expenditure of local government ≥ B then The economic development level =2	6	100%
If Area GDP ≥ D and Budgetary expenditure of local government ≥ B then The economic development level =3	1	100%
If Area GDP <D and Budgetary expenditure of local government <B then The economic development level =1	10	100%

As can be found from table 5, the probabilistic decision rules, as a result of deduction, has been generated by adopting the hybrid grey clustering and fuzzy rough sets reduction method. The confidence coefficient was 100% respectively. In addition, for all 17 regions that are correctly classified, the classification qualities of them are 100%. Hybrid model of grey clustering and fuzzy rough set method not only give the relative advantages of categories of the regional economic development level, but obtains probabilistic decision rules which set to minimum, providing a reasonable classification and overall evaluation of the level of regional economic development.

4 Conclusions


This paper is applied to evaluate process of regional economic in the decision-making in accordance with the grey clustering and rough set theory. The results are satisfactory. The generated preference model could deal with many uncertain incompatibilities appearing in the decision table with fewer rules. Due to the preferential model is “if Then” which is a specific preferential probabilistic decision rules. The established preferential model has been closed to the natural inference of the decision makers and could be understand easily. These rules not only explain the preference of decision makers, but also provide decision suggestions.




Acknowledgements

The authors would like to appreciate the financial support from the National Natural Science Fund Project 71173104; the major project supported by National Social Science 10zd&014; the Ministry of education of Humanities and Social Science Fund Project 09YJA630067; the operating mode of Jiangsu College of integrated social service 2012ZDIXM030; Research Institute of industrial technology of Jiangsu BR2012080; Industry-Academia-Research Project of Nanjing University of Aeronautics and Astronautics NC2012006; Project supported by “the Fundamental Research Funds for the Central Universities” NJ20130020; Humanities and Social Sciences youth project of the Ministry of education 12YJC630079; Social science planning research key project of Shandong 13BGLJ12.

References

- [1] Song marlin 2006 Economic development evaluation around the city in anhui province *Statistics Education* 79(4) 55-7
- [2] Luan Jinchang, Weng Jilin, Chi Guotai 2008 Economic development evaluation and decision based on line ahp *Statistics and decision* 271(19) 70-2
- [3] Wang Xiaoliang, Wang Ying 2013 Study on the construction of the evaluation index system of regional open economy’s development level *Research and development of the region* 32(3) 27-31
- [4] Sheng Minglan 2008 Suggestions to the evaluation and analysis of county economy development level-chongqing as an example *Journal of Southwestern Normal University* 33(6) 106-11
- [5] Wang Xizhao, Cui Aiping, Jin Hao 2011 Evaluation system and empirical analysis of regional competitiveness of dairy industry in north china *Journal of Hebei University of Economics and Business* 32(6) 90-6
- [6] Chen Yihua 2007 Journal of multilevel grey evaluation model and its application to the regional influence of chongqing *Journal of Chongqing University* 30(12) 141-5
- [7] Jian Lirong, Liu Sifeng, Xie Naiming 2010 Probabilistic decision method for hybrid grey clustering and extended dominance rough set *Journal of systems engineering* 25(4) 554-60
- [8] Greco S, Matarazzo B, Slow inskiR 2001 Rough sets theory for multicriteria decision analysis *European Journal of Operational Research* 129(1) 1-47
- [9] Liu Yong, Jian Lirong, Liu Sifeng 2012 Probabilistic decision method for hybrid grey clustering and variable precision rough fuzzy sets *Systems engineering* 30(5) 89-95
- [10] Pawlak Z, Skowron A 2007 Rough sets and Boolean reasoning *Information Science* 177(1) 41-73

Authors	
	<p>Li Peizhe, born in July, 1981, Binzhou County, Shandong Province, P.R. China</p> <p>Current position, grades: PHD Candidate of College of Economics and Management, Nanjing University of Aeronautics and Astronautics, China. Lecturer, Shan Dong University of Political science and Law, China.</p> <p>University studies: He received his M.Sc. from Shandong University of Science and Technology in China.</p> <p>Scientific interest: His research interest fields include Technology innovation management, Forecasting and decision-making.</p> <p>Publications: 10 papers published in various journals.</p> <p>Experience: He has teaching experience of 7 years, has participated ten scientific research projects.</p>

	<p>Jian Lirong, born in November, 1968, Jining County, Inner Mongolia, P.R. China</p> <p>Current position, grades: the Professor of School of College of Economics and Management, Nanjing University of Aeronautics and Astronautics, China.</p> <p>University studies: She received Her PHD in Management science and Engineering from Southeast University in China.</p> <p>Scientific interest: Her research interest fields include Innovation management, Management forecast and decision theory.</p> <p>Publications: 60 papers published in various journals.</p> <p>Experience: She has teaching experience of 20 years, has participated twenty scientific research projects.</p>
	<p>Zhang Kun, born in January, 1981, Jinan County, Shandong Province, P.R. China</p> <p>Current position, grades: PHD Candidate of College of Economics and Management, Nanjing University of Aeronautics and Astronautics, China. Lecturer, Shan Dong University of Political science and Law, China.</p> <p>University studies: He received his M.Sc. from Shandong University of Science and Technology in China.</p> <p>Scientific interest: His research interest fields include Environmental governance, Economic decision-making.</p> <p>Publications: 6 papers published in various journals.</p> <p>Experience: He has teaching experience of 7 years, has participated five scientific research projects.</p>
	<p>Pei Shanshan, born in October, 1984, Dongying County, Shandong Province, P.R. China</p> <p>Current position, grades: Lecturer, Shan Dong University of Political science and Law, China.</p> <p>University studies: She received his M.Sc. from Shandong University of Science and Technology in China.</p> <p>Scientific interest: Her research interest fields include Soft computing, data mining.</p> <p>Publications: 6 papers published in various journals.</p> <p>Experience: She has teaching experience of 7 years, has participated six scientific research projects.</p>

Power purchase plan using minimal cost flow

Zhan'an Zhang*, Xingguo Cai

School of Electrical Engineering and Automation, Harbin Institute of Technology, No.92, West Straight Street Harbin, Heilongjiang, China

Received 1 March 2014, www.tsi.lv

Abstract

The minimal cost flow method presented in this paper is used to calculate the minimum power purchase cost, because the user's goal is the minimal power purchasing cost under open electricity market. By establishing the minimal power, purchasing cost model and using the minimum cost flow algorithm to solve the objective function, the method is used to calculate a network in a single time. A continuous period of power purchase plan is obtained by accumulating each period network flow together. Example analysis proves that it is feasible to solve the problem with the minimal cost flow algorithm.

Keywords: Power Purchase Cost, Minimal Cost Flow, Network, Electric Market, Power Purchase Plan

1 Introduction

Before 2002, China's electric power system was a vertically integrated monopoly industry, with power plants and grids integrated within one enterprise. With the reform of the power system in China, the electricity market gradually opens to the users. The users can choose their own power supply companies. Direct-purchasing electricity was first performed on large customers [1-3], and then gradually extended to all users. The user's goal is to take minimum power purchase cost as a power purchase plan. Therefore, the minimal purchase cost research has very practical significance.

Different countries have different power systems hence electricity-purchasing researches are different [4-6]. The purpose of China's electric power system reform is to break the monopoly and form competition. This is helpful to power plants, which need to improve efficiency and to power grid in terms of the reduction of line losses, which is beneficial to the development of the power system.

Current electricity purchasing cost model generally does not consider the structure of the grid. As the network flow algorithm can reserve system network topology structure and has the advantage of easiness in considering security constraints, it is suitable for solving constraints of high dimension, linear and nonlinear optimization problems. Therefore, the network flow theory has been widely applied to many fields such as power system dispatching, power system restoration plans [7], observability analysis [8, 9], network reliability [10-13], etc.

In this paper, the minimum cost flow algorithm is used to solve the minimal purchasing cost model and calculate a network in a single time. A continuous period of power purchase plan is obtained by accumulating each

period network flow together. Case studies have proven that the minimal cost flow method was successful in solving the minimal purchase cost problem and therefore has a certain applicational value.

2 The mathematical model

2.1 THE OBJECTIVE FUNCTION

The user's goal is the minimal electricity purchasing cost under open electricity market. The objective function can be expressed as:

$$\min \sum_{t=1}^T \sum_{i=1}^N W_{i,t} \rho_{i,t}, \quad (1)$$

$$W_{i,t} = P_{i,t} \Delta t, \quad (2)$$

where $W_{i,t}$ is the electricity of the i th branch at t time; $\rho_{i,t}$ is the cost of unit electricity corresponding to $W_{i,t}$; $P_{i,t}$ is the power of the i th branch at t time; N represents the number of branches.

2.2 CONSTRAINTS

The objective function need to satisfy the following constraints:

(1) Power constraints

The power constraints include the generator output and line transmission limit.

$$P_{i\min} \leq P_{i,t} \leq P_{i\max} \quad i = 1, 2, \dots, N, \quad (3)$$

where $P_{i\min}$ is the lower and $P_{i\max}$ is the upper output limit of unit i , respectively.

(2) Node voltage constraints

* Corresponding author - Tel: +86-152-046-27420 fax: +86-451-536-81710; E-mail: zza.163.com@126.com

$$U_{i\min} \leq U_{i,t} \leq U_{i\max} \quad i = 1, 2, \dots, N, \quad (4)$$

where $U_{i,t}$ is the voltage of node i at t time; $U_{i\min}$ is the lower and $U_{i\max}$ is the upper limit of node i , respectively.

(3) Power balance constraints

$$\sum_{i=1}^N P_{i,t} = L_t \quad i = 1, 2, \dots, N, \quad (5)$$

where L_t is the system load at t time.

3 Network flow theory introduction

3.1 NETWORK FLOW INTRODUCTION

Network flow problems belong to the category of graph theory. Graph is also referred to as a network, consisting of a number of nodes and its connecting arcs. If any two nodes in graph G are connected, then G is a connected graph. It is called a connected digraph if each arc has a direction. For digraph $G = (V, E)$, where V is the set of all nodes; E is the set of all arcs. The node which only has outflow arcs is called source point(s), while the one only has inflow arcs is called receive point(r) in V . The rest nodes which there are both outflow and inflow arcs are called intermediate points.

Arc(i,j) is the edge connecting node i and j , which has a predetermined direction, flow f_{ij} , the cost of the unit flow c_{ij} , the flow rate limit (as shown in Figure1 and Figure 2).

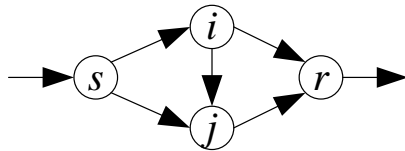


FIGURE 1 A Network

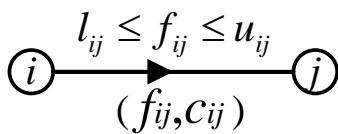


FIGURE 2 Flow Limit and Cost in An Arc

A set of f that meets the following two basic conditions in the graph can be viewed as a feasible flow.

(1) For each arc to meet capacity limits.

$$l_{ij} \leq f_{ij} \leq u_{ij} \quad (i, j) \in E. \quad (6)$$

where l_{ij} is lower limit, u_{ij} is upper limit.

(2) For the intermediate node, inflow is equal to outflow.

$$\sum_{(i,j) \in E} f_{ij} = \sum_{(k,i) \in E} f_{ki} \quad (7)$$

It can be proved that the net flows are equal from s to r . The flows are all set to v , and v is called the flow of the feasible flow f [14].

3.2 THE MINIMAL COST FLOW METHOD

The minimum cost flow problem is to make the minimum total cost under the specified flow taking into account the arc cost. The objective function can be expressed as:

$$\min \sum_{(i,j) \in E} c_{ij} f_{ij} \quad (8)$$

The minimal cost flow algorithm is mainly Klein algorithm, scaling algorithm, out-of-kilter and minimum cost path algorithm, etc. We take Busacker-Gowan iteration method here [15], which is relatively mature, the main steps are as follows:

- (1) Find out minimum cost path μ from s to r ;
- (2) Distribute the maximum possible flow f to μ :

$$f = \min \sum_{(i,j) \in \mu} \{u_{ij}\} \quad (9)$$

And let all edges capacity reduce f . In this case, the unit flow cost should be changed to infinity for the saturated edges;

- (3) Make the reverse side of all edge (i,j) on the pathway μ , then $u_{ji} = f, c_{ji} = -c_{ij}$;
- (4) In such a new network, repeat the above steps (1), (2), (3) until the complete flow is equal to v .

3.3 MULTI-PERIOD NETWORK FLOW MODEL

A continuous time network flows is constituted by combining each network flow chart of a single time in space. Multi-period model is a three-dimensional model with each layer corresponding to a single period model (as shown in Figure 3). Each s_i is united together to form new s , in the same way to form new r . Arc(s, s_i) represents the output, and arc(r_i, r) represents the load. Considering each period together, the power purchase plan of continuous time can be obtained.

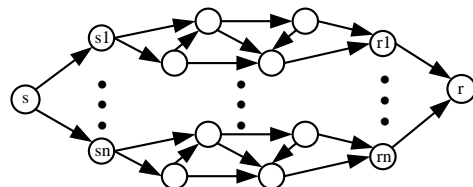


FIGURE 3 Multi-Period Network Flow Model

The above structure model is a typical generalized network flow programming problem. It can be expressed as:

$$\min C^T F, \quad (10)$$

$$\text{satisfied with } AF = b, \quad (11)$$

where C represents the m -dimensional arc cost vector; F represents the m -dimensional arc flow; b represents an n -dimensional vector injected nodes; A represents $n \times m$ dimensional node-arc incidence matrix [16].

4 Example analysis

A system diagram and its corresponding cost flow network diagram are shown in Figure 4 and Figure 5 respectively. When the purchasing electricity changes from minimum to maximum, we calculate the corresponding changes in the minimal power purchase costs. The unit of arc flow and its unit cost are MWh and Yuan/MWh respectively. For example, (15,330) means that the upper flow is 15 MWh and the unit cost is 330 Yuan/MWh.

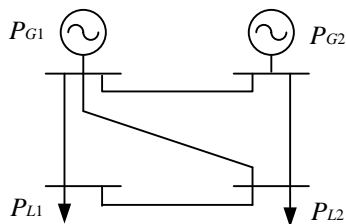


FIGURE 4 Power System Diagram

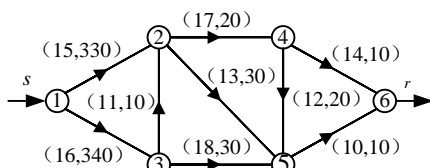


FIGURE 5 Network Diagram

Arc capacity matrix F and arc cost matrix C are

$$F = \begin{bmatrix} 0 & 15 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 17 & 13 & 0 \\ 0 & 11 & 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 12 & 14 \\ 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 330 & 340 & 0 & 0 & 0 \\ 0 & 0 & 0 & 20 & 30 & 0 \\ 0 & 10 & 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 0 & 20 & 10 \\ 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Put F and C into the program, the largest flow is 24MWh, the corresponding minimum cost is 8740Yuan, and the minimal cost matrix F_{min} is

$$F_{min} = \begin{bmatrix} 0 & 15 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 14 & 1 & 0 \\ 0 & 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 14 \\ 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Minimum cost flow calculation is shown in Table 1 by the maximum flow[17], where C is the unit purchase cost, F is the purchase electricity, and Q is the electricity purchase cost.

TABLE 1 The minimum cost flow calculation

Step	Path	C(Yuan/MWh)	F(MWh)	Q(Yuan)
1	1→2→4→6	360	14	5040
2	1→2→5→6	370	1	370
3	1→3→5→6	370	9	3330

The distribution of the minimum cost flow can be divided into several steps to solve when the maximum flow passed through the network. Combining the steps together, the final result of the minimum cost flow distribution can be obtained (as shown in Figure 6 and Figure 7).

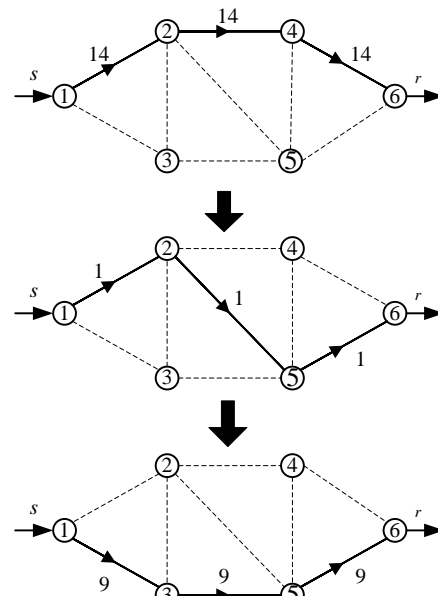


FIGURE 6 The Minimal Cost Flow Solving Process

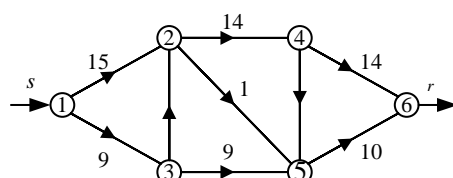


FIGURE 7 The Minimal Cost Flow Distribution

Assuming that the system is running in the same way, when the flow changes from minimum to maximum, the corresponding minimum costs will also increase (as shown in Table 2).

TABLE 2 The minimum cost under different flow

F	1	2	3	4	5	6	7	8
Q	360	720	1080	1440	1800	2160	2520	2880
F	9	10	11	12	13	14	15	16
Q	3240	3600	3960	4320	4680	5040	5410	5780
F	17	18	19	20	21	22	23	24
Q	6150	6520	6890	7260	7630	8000	8370	8740

Assuming that the system is running in period T , the purchasing power and the corresponding minimal cost change with the same trend(as shown in Table 3 and Figure 8). Accumulated purchasing power will gradually increase with the increase of time, the same trend with the accumulated minimal power purchase cost (as shown in Figure 9). Then the optimal power purchase scheme can be obtained from the above curves, which can clearly reflect the power-purchasing path.

TABLE 3 Each period flow distribution

T	1	2	3	4	5	6	7	8
F	18	17	16	17	18	19	20	21
Q	6520	6150	5780	6150	6520	6890	7260	7630
T	9	10	11	12	13	14	15	16
F	22	23	22	20	18	19	20	21
Q	8000	8370	8000	7260	6520	6890	7260	7630
T	17	18	19	20	21	22	23	24
F	22	23	24	23	22	21	20	19
Q	8000	8370	8740	8370	8000	7630	7260	6890

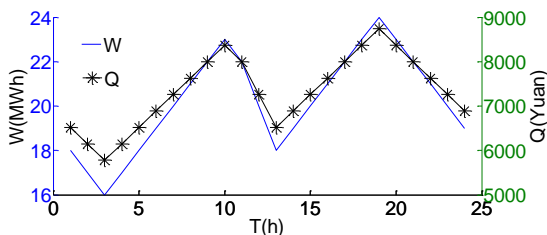


FIGURE 8 Purchasing Power and Minimal Power Purchase Cost Curve

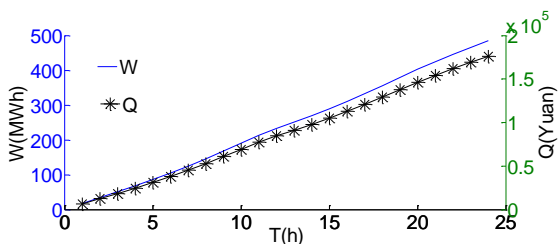


FIGURE 9 Purchasing Power and Minimal Power Purchase Cost Accumulation Curve

When the system operation mode changes, by changing the arc cost matrix and the arc capacity matrix, we can easily adapt to change in operating mode.

5 Conclusions

The minimal cost flow algorithm is used to determine the power purchase plan, which can reflect the system network topology making it easy to consider system constraints. The method is simple, rapid and clear in physical concept. When the operation mode changes, by



changing the network matrix we can quickly adapt to the new operation mode. Example analysis proves that the method is feasible and practical.

Acknowledgments

In the process of writing paper, thanks for the help and guidance of my colleagues and supervisor.

References

- [1] Jian S, Yanling W 2009 Research on the admission standard of large consumers direct-purchasing based on the externality theory *Power and Energy Engineering Conference* 1-5
- [2] Yanan Z, Gengyin L, Ming Z, Xuanxuan L 2008 An integrated approach on allocating the congestion management cost under large consumers electricity direct purchase mode *The 3th International Conference on deregulation and restructuring and Power Technologies* 539-44
- [3] Hongchun S, Fang Z, Jun D, Xu H 2008 Research on transmission price making mechanism of direct power purchasing for large electric power consumers in Yunnan province *The 3th International Conference on deregulation and restructuring and Power Technologies* 463-7
- [4] Bhanot K 2002 Value of an option to purchase electric power-the case of uncertain consumption *Energy Economics* 24(2) 121-37
- [5] Lan Z, Luh P, Xiaohong G, Merchel G 1994 Optimizaion-based inter-utility power purchases *IEEE Transactions on Power System* 9(2) 891-7
- [6] Debin F, Jingfang W, Dawei T 2012 A double auction for competitive generators and large consumers considering power transmission cost *Electrical Power and Energy Systems* 43(1) 880-8
- [7] Mota A, Mota L, Morelato A 2007 Visulization of power system restoration plans using CPM/PERT graphs *IEEE Transactions on Power System* 22(3) 1322-9
- [8] George N, Peter J 2003 A hybrid method for observability analysis using a reduced network graph theory *IEEE Transactions on Power System* 18(1) 295-304
- [9] George N, Peter J, Kevin A, Davis P 2003 Numerical observability analysis based on network graph theory *IEEE Transactions on Power System* 18(3) 1035-45
- [10] Weichang Y 2008 A simple minimal path method for estimating the weighted multi-commodity multistate unreliable networks reliability *Reliability Engineering and System Safety* 93(1) 125-36
- [11] Yi-Kuei L 2004 Reliability of a stochastic-flow network with unreliable branches & nodes, under budget constraints *IEEE Transactions on Reliability* 53(3) 381-7
- [12] Yi-Kuei L 2007 Reliability of a flow network subject to budget constraints *IEEE Transactions on Reliability* 56(1) 10-6
- [13] Yi-Kuei L 2007 System reliability of a limited-flow network in multicommodity case *IEEE Transactions on Reliability* 56(1) 17-25
- [14] Jianxue W, Xifan W, Xiuli W 2008 Regional reserve model based on network flow theory *Proceedings of the CSEE* 28(10) 51-5 (in Chinese)
- [15] Suixiang G 2009 *Graph theory and network flow theory* Beijing. Press: Higher Education Press (in Chinese)
- [16] Erkeng Y, Guangyi L, Jingyang Z 2001 *Energy management system* Beijing. Press: Science Press (in Chinese)
- [17] Jishou L, Chuan Z 1993 *Optimal operation of the power system* Wuhan. Press: Huazhong University of Science and Technology Press (in Chinese)

Authors	
	<p>Zhan'an Zhang, born in February, 1975, Harbin City, Heilongjiang Province, P.R. China</p> <p>Current position, grades: doctoral candidate of Electrical Engineering, Harbin Institute of Technology. University studies: Graduated from Northeast Dianli University in 2009, received a master's degree in Electrical Engineering. Scientific interest: Power System and Its Economic Dispatch. Publications: more than 10 papers published in various journals.</p>
	<p>Xingguo Cai, born in August 1945, Harbin City, Heilongjiang Province, P.R. China</p> <p>Current position, grades: professor, doctoral supervisor, Harbin Institute of Technology. University studies: Graduated from Harbin Institute of Technology in 1983, received a master's degree in Electrical Engineering. Scientific interest: Power System and Its Economic Dispatch. Publications: Presided over the completion of national and provincial, more than 10 scientific research projects; more than 50 papers published in various journals.</p>

Novel approaches to form the corporate culture in pedagogical universities

T Koycheva*

South Ukrainian National Pedagogical University named after K D Ushinsky

Ukraine, Odessa, Staroportofrankovskaya 26, 65020

Received 25 October 2013, www.tsi.lv

Abstract

The role of a corporate culture in the development of the modern education, especially in the improvement of the university education is considered. On the example of pedagogical university it's shown that the active research work promotes the formation of a corporate culture within the educational institution. At the same time an availability of the corporate culture leads to the expansion and intensification of scientific work. As a result of our research it was confirmed a particular importance of scientific schools and international consortiums in the formation of a corporate culture.

Keywords: corporate culture, pedagogical universities, research work, scientific schools

1 Introduction

The education system in each country is continually evolving and improving. From time to time new reforms and innovations are proposed. Their effectiveness is tested by practice. The relationship of reforms to innovations requires a special consideration. Here it is only worth mentioning that a reform can be described as a 'top-down' approach. In this case the initiative of a team is limited. At the same time the innovation is characterized as a "bottom-up" or "grassroots" approach, and different individuals and organizations can show their creativity. Besides that we can speak about reforms in the conventional sense (explicit reforms) and implicit reforms that just are introducing innovations.

In educational institutions technological improvements are also practiced (different types of schedules, different types of knowledge control, etc.). Technology has a certain value and can improve a higher education as it happens in other knowledge-based sectors like music, journalism, and financial services, where new providers improve the access and convenience while reducing costs. It is clear that technology does not guarantee innovation. The prerogative of each educational institution is to determine the processes by which innovations emerge and create own innovative approaches. It should be noted that many innovative practices have no scholarly literature base. We investigated *the ways of a creation of own corporate culture* in the pedagogical university as an effective innovation approach to improve the educational process.

A corporate culture is a significant factor for prosperity of a company, any institution or university.

The modern companies pay special attention to formation of their corporate culture. This trend is typical also for the education sector where important social problems such as a training of specialists, educational services, information processing etc. are solving. There is a direct connection between the employees and the means of production. This relationship determines a character of the necessary corporate culture. We can find in publications different definitions and various schemes for the corporate culture. For example, [1] in Figure 1 four basic types of corporate culture are shown. Usually these types are combined dependently on the problems that an organization should solve. The appropriate activities that could lead to formation of different types of corporate culture can be carried out on any level beginning from authorities and to ordinary performers. Any initiative should be taken by a whole team. Further experience confirms the feasibility of this initiative.

As to educational institution it is naturally to choose "Incubator type" and "Guided missile" types as the main strategy of corporation culture formation. It is extremely important that a corporate culture evolved strategically in order to optimize the development of institution. For an educational institution such as university the introduction of innovations is possible first of all through the use of internal resources of the university, of which the most important is the intellectual potential of the collective [2]. This internal resource contributes significantly to the formation of the corporate culture of an educational institution and to the progress in the educational process generally.

* Corresponding author - Tel. 38048 7325107,
E-mail: koycheva@mail.ru

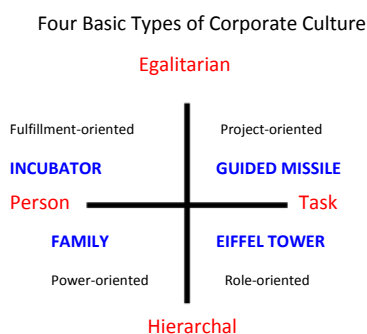


FIGURE 1 One of possible schemes for corporate culture

In the XXI century the main directions of the university education improvement are determined as:

- ✓ Educational, which is embodied in the transfer of knowledge from one generation of specialists to another;
- ✓ Exploratory, that is provided by the development of science;
- ✓ Servicing, as the service of the state and society;
- ✓ Cultural, that ensures the development of culture and intelligence;
- ✓ Humanistic, as a broadcast of important ideas that determine the sense of human life and also its activities [3].

The corporate culture of an educational institution should be formed on the base of these main directions of the university education improvement adopted at the present time. The psychological and moral climate of university, the ability of its employees to act together in line with common values and interests is understood today as corporate culture. The modern university is a socio-cultural institution where not only finished forms of culture are broadcast mainly in the form of knowledge but also where these forms are created and where one learns how to create them. Corporate culture determines the place of each university in the world embodying the unwritten laws, rules and regulations that unite its employees and link them together in common activities.

2 Corporate culture in pedagogical university

The problem of creating a corporate culture in pedagogical university has its own features. First, it is necessary to take into account the general cultural traditions of the country. They are formed in the course of the historical process and can be linked to mental specificity of the population. It is of great educational value. For example, in order to illustrate the mental specificity of different countries, in Table 1 the specificity of the relationship of employer and employee in USA and Japan is presented.

TABLE 1 The features of the relationship of employer and employee in USA and Japan [4]

USA	Japan
Do not value emotional sensitivity	Valued emotional sensitivity
Straight-forward, impersonal	Hide emotions
Litigation; little conciliation	Conciliations
Low commitment to employer; willing to break ties if needed	Loyal to employer; employer takes care of employees; extreme reluctance to break ties
Individual leader decision making (with input from team)	Group decision by consensus
Methodical decision making	Step-by-step decision making
Profit or good of individual is ultimate aim	Good of group is highest aim

Defining the specific conditions of formation of corporate culture in the pedagogical universities, we proceeded from the fact that the starting point in this process should be the mission of institution. Everything that has an impact on the corporate culture (choice of strategies, management style, a personality of the leader, common values, and traditions) is caused by the mission of the institution. The mission of any enterprise is its vital goal [5].

The effective corporate culture assumes:

- ✓ Teamwork, cooperation (team spirit);
- ✓ Job satisfaction and the pride for the results;
- ✓ Dedication to the organization and willingness to meet its high standards;
- ✓ High demands on the quality of work;
- ✓ Readiness for change, due to the requirement of progress and competition, despite the difficulties and bureaucratic obstacles.

A pedagogical university solves its own specific tasks on the base of pedagogical and psychological approaches. We define such characteristic features of corporate culture in pedagogical university:

- ✓ Highly favourable moral climate;
- ✓ A high level of ethical relations;
- ✓ A high motivation of teachers to achieve the best educational effect;
- ✓ A deep link of scientific research with the educational process, especially in the fields of psychology and pedagogy;
- ✓ A high level of responsibility and organization.

It is not simple to define the realistic 'production' mission for the pedagogical university, which is under the strong influence of the state educational policy and also mental features and values of Soviet period. Therefore, we turned to the field of the scientific research, which is a specific task for the teachers of any university and in particular of pedagogical university. Organization of scientific research largely determines the features of corporate culture in pedagogical university.

3 The ways of formation of corporate culture in pedagogical university

Similarly to large enterprises that combine different types of production activities, universities and especially pedagogical universities can be defined as special corporations, where obvious and traditional educational activities are linked to scientific, methodological, organizational and other types of socially important activities [6]. On the basis of this provision we consider the efficiency of pedagogical university operation in terms of optimal management of scientific research.

An active research work influences the efficiency of the university activities on the whole in two directions: an increase of the educational level and a formation of the appropriate corporate culture.

An impact of a scientific research on the educational process is reflected in:

- ✓ Enriching the educational content with new scientific ideas and the development of new educational standards;
- ✓ Improvement of techniques and methodologies of teaching process on the basis of research;
- ✓ Research in the pedagogical field as an effective way to improve the professional skills of teachers;
- ✓ Development of creative abilities and cognitive activity of students;
- ✓ Determining the ways and prospects for the university education.

As to link of scientific research to formation of corporate culture, two questions have to be clarified. The first question is what features of the organization and performance of scientific work can significantly influence a corporate culture of pedagogical university. The second question is how to use optimally the intellectual potential of teaching staff to achieve a high level of scientific research and a maximal effect in the improvement of all activities of the university through formation of the appropriate corporate culture.

We claim that the scientific research in pedagogical university provides an atmosphere of creative students' attitudes toward their future professional occupation. It is an important element of a corporate culture of pedagogical university. A special influence on the formation of corporate culture has reputable scientific schools. Our experimental results confirm this statement.

A scientific school is a special form of collaboration of scientists in the course of development of science. In relation to pedagogical university, a scientific school is linked to the goals and objectives of education, the system of pedagogical values, the basic methodological ideas, etc.

We select such features of scientific school [7]:

- ✓ Recognition of the leadership of the founder of scientific school;
- ✓ Common ideological views and common approaches to solve scientific problems;

- ✓ Common ethical and moral principles;
- ✓ Continuity in solving problems from one generation to another;
- ✓ Close connections with the related scientific schools;
- ✓ Objective evaluation of scientific results of own and related scientific schools.

4 Results and discussion

We studied a development of scientific schools in South-Ukrainian national pedagogical university after K. D. Ushinsky and their influence on the dynamics of a corporate culture of the collective. During five years, we observed how thanks to efforts of the authority the role of scientific schools were increased. The results of research activities of leading scientists and scientific schools have been widely promoted in the educational process, covered in the university press. The most important scientific results were discussed at numerous seminars, meetings, at the university scientific council, and at conversations with students [8].

It is necessary to highlight the role of interdisciplinary research in the formation of corporate culture in university. Strong links between scientific schools of different faculties are formed.

Table 2 shows the results concerning the distribution of teachers by the types of corporate culture in the beginning of our research. All teachers have PhD degree. The method of Cameron-Quinn [9] was used. The method consists of two basic elements: a culture model and an assessment instrument. Thanks to the method's consistent logic, simplicity, and scientific soundness, it can be implemented in the cultural discussion without a lengthy training period.

The Cameron-Quinn method operates on the principle that a particular culture is neither good nor bad per se. What matters most is whether the culture fits with the organization and its strategy. In contrast to many other culture models, it is also based on the assumption that culture cannot be grown arbitrarily. In other words, if a culture characteristic needs to be reinforced, it is always necessary to clarify which other characteristic has to be reduced. This discussion unequivocally leads to a new level of precision and relevance

As it is seen in Table 2, for a group of 134 teachers with the scientific degree of PhD the most teachers belong to a hierarchical type of corporate culture (28.6%). Fewer teachers revealed their penchant for clannish (25.2%) and market (24.0%) type of corporate culture. The last part revealed a commitment to adhocracy type of corporate culture (22.2%). This trend continues regardless of the age of teachers who participated in the survey and the effectiveness of their research activities. One can see that the differences in the belonging to considered types of corporate culture are small, between 22.2 - 28.6%.

TABLE 2 Distribution of teachers between different types of corporate culture

The range of scores for science activity	Average score	Number of teachers	Average age	Types of culture			
				Clannish	Adhocracy	Market	Hierarchical
100-999	603,8	48	45,1	26,4	23,4	24,3	25,9
1000-1999	1425,2	41	45,2	25,5	22,1	23,5	2,9
2000-2999	2381,9	26	45,7	24,4	21,7	24,	29,6
3000-3999	3523,8	13	43,2	25,3	21,8	22,7	30,2
More than 4000	5065,0	6	50,7	18,5	20,4	27,1	34,0
In the total sample 0-8000	1635,1	134	46	25,2	22,2	24,0	28,6

This tendency corresponds to a structure of the corporate culture of international scientific schools [10]. Our observations showed that a participation of scientists in international consortiums influences on the formation of a corporate culture in university, especially on its structure. We observed such chain of linked processes: a formation and structuring of scientific schools – increase of the number of students involved in the creative work – a formation of a specific corporate culture in university – increase of the overall educational level in university.

5 Conclusions

The main features of the corporate culture in pedagogical university are determined. The ways of a formation of the appropriate corporate culture are analysed. The role of scientific schools in the formation and structuring of a corporate culture is studied. Significant factors that influence on the formation of a corporate culture in pedagogical university are revealed. A relationship between the functioning of scientific schools, the features of a corporate culture in the university and the overall educational level is established.

Acknowledgments

Author would like to acknowledge useful discussions with Prof. Ela Karpova and Prof. Arnold Kiv.

References

- [1] Volvach V G 2010 *Information education and economics* 6(17) 100-107 (in Russian)
- [2] Spivak V A 2001 *Korporativnaya kultura* SPB: Piter 352 p (in Russian)
- [3] Gregory K 1983 *Administrative Science Quarterly* 28(2) 359-76
- [4] Trompenaars F, Hampden-Turner H 1997 *The Seven Dimensions of Culture* <http://www.mindtools.com/pages/article/seven-dimensions.htm> (16 Dec 2012)
- [5] Koycheva T I 2012 *Visnik Cherkas'kogo universitetu (Bulletin of Cherkassy university)* 28(241) 49-55 (in Ukrainian)
- [6] Koycheva T I 2011 The role of scientific research in the innovation development of universities *The modern information technologies and innovation methods* ed V I Bykov Planer: Kiev 362-7 (in Russian)
- [7] Yablonskene N L 2006 *Korporativnaya kultura sovremenogo universiteta Universitetskoe upravlenie (University management)* No 2 8-15 <http://ksp-ed.union.ru/doc/yablonskene.pdf> (in Russian)
- [8] Koycheva T I 2013 Scientific schools as a source of development of a corporate culture of pedagogical university *Pedagogy of higher and secondary school* 37(1) 240-245 (in Russian)
- [9] *National Center for Postsecondary Improvement* ed J F Greden. <http://ncpi.stanford.edu> (20 Oct 2013)
- [10] Jones S M and Bouffard S M 2012 Social and Emotional Learning in Schools: From Programs to Strategies *Social Policy Report* Society for Research in Child Development www.srcd.org (18 Apr 2013)

Author



Tatyana Koycheva

Current position, grades:

PhD in pedagogical sciences, As. Professor of the Department of Informatics and Applied Mathematics in South-Ukrainian National Pedagogical University after K. D. Ushinskij.

Scientific interest: innovative education, distance education, influence of scientific communities on the corporate culture formation.

Experience: Head of Scientific-Organizational department communities on the corporate culture formation.



Authors' index					
Aitmagambetov Altay	166	Li Longshu	250	Wang Chengji	126
Ao Jianfeng	151	Li Peizhe	274	Wang Hongbao	259
Bellucci Stefano	7	Li Wei	27	Wang Jie	180
Bi Jianxin	240	Li Zhengming	95	Wang Mei	57
Burlutskaya Nataly	7	Liang Xiao	27	Wang Peigang	143
Butuzov Yuri	166	Ling Chen	95	Wang Qin	135
Cai Xingguo	281	Liu Chunxiao	40	Wang Shaoren	265
Caihong Liu	212	Liu Guixian	70	Wei Xingping	52
Capobianchi Aldo	7	Liu Jianxiao	126	Wei Xiong	212
Chen Hua	33	Liu Ranran	95	Wu Jie	33
Chen Jianping	232	Liu Xiaoyong	76	Xiang Liuming	161
Chen Qing	161	Liu Xiaoyu	52	Xiu Baoxin	192
Cheng Guozhu	40	Liu Yang	57	Xu Yang	70
Cong Guodong	221	Liu Zhi-Hui	57	Yang Guoyan	114
Duan Kanghong	143	Liu Zhong	192	Yang Xianhui	107
Ershkovich A	24	Liu Zhusheng	205	Yang Xuemei	199
Fan Weitong	27	Lobanova-Shunina Tamara	7	Yang Yongchong	157
Feng Yanghe	192	Lou Ning	157	Yao Sheng	250
Gao Yu-Xiao	57	Ma Zhihong	232	Yue Qi	135
Gong Yusheng	83	Ma Zujun	265	Yuejing Lv	46
Gopejenko Victor	7	Meng Xian-Yong	57	Yulong Pei	186
Gu Lichuan	126	Micciulla Federico	7	Zeng Shouzhen	232
Guan Xin	114	Mukazhanov Nurzhan	170	Zhang Hong	40
Guozhu Cheng	186	Nahla Mohamed Ahmed	95	Zhang Hongxin	143
Hong Zhang	46	Pei Shanshan	274	Zhang Jiong	70
Jian Lirong	274	Ren Honge	107	Zhang Kun	274
Jiao Junsheng	63	Shang Wei	180	Zhang Yaping	40
Jie Li	46	Shi Shangyao	175	Zhang Yi	76
Jin Maozhu	180	Shi Shaobo	135	Zhang Zhan'an	281
Jing Guoqing	70	Shimei Wu	186	Zhao Guocheng	27
Jing Weipeng	107	Shunin Yuri	7	Zhao Shiping	52
Koycheva Tatyana	286	Song Bao-wei	33	Zhao Shiping	175
Kurmangaliyeva Bikesh	170	Song Shilin	143	Zhao Suxia	199
Lei Lianghai	240	Song Xiao-ru	33	Zheng Xiaobo	157
Li Bingjie	227	Su Tiexiong	76	Zhong Yong	161
Li Huang	99	Sun Peihong	199	Zhuang Bochao	265
Li Huisong	120	Torekhan S	166	Zhukovskii Yuri	7
Li Jun	52, 175	Uskenbayeva Raissa	170		

Cumulative Index

Nanoscience and Nanotechnology

Yu N Shunin, S Bellucci, Yu F Zhukovskii, V I Gopejenko, N Burlutskaya, T Lobanova-Shunina, A Capobianchi, F Micciulla CNT-Fe-Pt interconnect electromagnetic simulations for magnetically stimulated CNT growth and novel memory nanodevices

Computer Modelling & New Technologies 2014 18(2) 7-23

The parametrically controlled production of carbon nanotubes (CNTs) with predefined morphologies is a topical technological problem for modern nanoelectronics. The chemical vapor deposition (CVD) technique for single walled carbon nanotubes (SWCNTs) in the presence of various metal nanoparticle catalysts is generally used now. The application of a magnetically stimulated CVD process scheme and catalyst nanoparticles with a strong magnetism promises additional possibilities for the CVD process management and allows expecting a predictable growth of CNTs with set chiralities and diameters. The main attention is focused on the magnetically anisotropy Pt-Fe in L10 crystallographic phase nanoparticles effect research. The developed cluster approach based on the multiple scattering theory formalism, realistic analytical and coherent potentials, as well as effective medium approximation (EMA-CPA), can be effectively used for modeling of nanosized systems. It allows us to calculate the dispersion law $E(k)$, electronic density of states, conductivity, etc. This theoretical approach is used for simulation of fundamental electromagnetic properties in Pt-Fe L10-CNT interconnects, which are responsible for developing CNTs morphologies. The developed model of 'effective bonds' and the model of magnetic stimulation for growing CNTs morphologies generated on the Pt-Fe nanoparticle surface are applied for the evaluation of the expected CNT chiralities distribution. The model and conditions controlled magnetically, which stimulate CNT growth in the CVD process, aimed at the predictable SWCNT diameter and chirality and based on Pt-Fe L10 catalyst are discussed. The possibilities of CNT forest growing on FePt nanoparticles for magnetic nanomemory are also evaluated.

Keywords: chemical vapor deposition (CVD), CNTs CVD magnetically controlled growth, arc discharge technique CNTs creation, Pt-Fe nanodrops –catalysts, fundamental electromagnetic properties in Pt-Fe L10-CNT interconnect, model of magnetically controlled CNTs growth

Fundamental Nature Phenomena

A Ershkovich On the classical roots of the Schroedinger equation

Computer Modelling & New Technologies 2014 18(2) 24-26

In semiclassical approximation Schroedinger equation is known to reduce to classical Hamilton-Jacobi equation. These equations look strikingly similar. An idea that just the Hamilton-Jacobi equation became a prototype of the Schroedinger equation arises. Arguments in favour of this assumption are supplied. Then it is no wonder that Aharonov-Bohm effect was recently derived directly from the classical Hamilton-Jacobi equation (without using Schroedinger equation), and hence, it is, in fact, of classical origin. The electron-field interaction is explained within the framework of classical electrodynamics. Thus, the so-called unlocal interaction becomes unnecessary.

Keywords: Schroedinger equation, Hamilton-Jacobi equation, Aharonov-Bohm effect

Mathematical and Computer Modelling

Xiao Liang, Wei Li, Weitong Fan, Guocheng Zhao Numerical simulation and hydrodynamic performance prediction for hydroplane longitudinal motion

Computer Modelling & New Technologies 2014 18(2) 27-32

Aiming at hydrodynamic performance prediction for hydroplane longitudinal motion, numerical simulation for a hydroplane motion was carried out by using VOF and RNG $k-\varepsilon$ model and solving Navier-Stokes equation under FLUENT software platform. Evolution of ship resistance was obtained as the velocity change, and flow field situation and dynamic pressure variation of hydroplane hull bottom were reflected intuitively. By comparing and analysing the results among numerical simulation calculation and ship model experiments and theoretical estimation, it was verified that hydrodynamic performance prediction for hydroplane longitudinal motion based on numerical simulation calculation under FLUENT is feasible and precise enough.

Keywords: numerical simulation, hydrodynamic performance prediction, hydroplane, FLUENT

Song Xiao-ru, Chen Hua, Song Bao-wei, Wu Jie Research on adaptive AUV tracking control system based on least squares support vector machine

Computer Modelling & New Technologies 2014 18(2) 33-39

Aimed at the nonlinear, uncertainties and impreciseness in AUV tracking control, an adaptive LSSVM control is highlighted. It is including the approximate dynamic inversion control law, linear controller and the robust adaptive LSSVM controller. The key is designed that adaptive LSSVM controller, which is to decrease or offset the uncertain dynamic inverse model errors. The online adjustment LSSVM parameters rules are deduced by the Lyapunov stability theorem. So the closed-loop tracking control system's stability and asymptotical convergence of tracking error can be guaranteed. It can be seen that the tracking errors converge and stay at a small neighbourhood of zero. At last taken a certain real AUV as an example, the numerical simulation results are given in the presence of the ocean current wave interference. They show the proposed method has good tracking performance and a certain robustness against modelling errors.

Keywords: AUV, LSSVM, tracking control, adaptive controller

Yaping Zhang, Chunxiao Liu, Hong Zhang, Guozhu Cheng A study on capacity of signalized intersection under snow-ice conditions based on classical model modification

Computer Modelling & New Technologies 2014 18(2) 40-45

As the common climatic phenomena frequently occurring in northern China in winter, snow and ice weathers have made great influence to the capacity of signalized intersection. By starting from snow and ice conditions, this paper makes improved correction about the calculation of capacity of signalized intersection respectively based on HCM Method and Method of Stopping Line. On the basis of HCM Method, the author comprehensively considers factors influencing the capacity of signalized intersection under snow and ice conditions, and calibrates the integrated parameters influencing the capacity of signalized intersection under snow and ice conditions; grounded on Method of Stopping Line, the author then calibrates various parameters and makes classified study about each parameter separately according to snow weather and freezing condition, finally giving out parameter values under various states. With the help of the revised model, the paper calculates the capacity of signalized intersection and then makes the comparison between traffic capacities under various snow and ice conditions and that with capacity of signalized intersection under normal weather, hence with strong practical significance.

Keywords: Snow and ice conditions; Signalized intersection; capacity; Parameter modification; Comparative analysis

Zhang Hong, Li Jie, Lv Yuejing Structure performance analysis of Vehicular Ad Hoc Networks based on complex network theory

Computer Modelling & New Technologies 2014 18(2) 46-51

A very promising direction in intelligent transportation system is the applications based on vehicular ad-hoc networks (VANET). VANET of a country is one of the most importation indicators of its economic growth. In this paper, we analysed the characteristics of VANET network topology using complex network theory. The author contribute VANET model, analyse the statistical properties of complex network based on the degree distribution, the clustering coefficient, the average path length and find that the network has scale-free and small world features. The structure and properties of VANET have great implications for traffic congestion and urban planning. The robustness analysis indicates that VANET is robust to random attack when considering static robustness, but somewhat vulnerable to intentional attack.

Keywords: complex network, VANET, topological analysis, robustness

Xingping Wei, Jun Li, Xiaoyu Liu, Shiping Zhao A fast on-line two-dimensional sizes measurement method for micro part

Computer Modelling & New Technologies 2014 18(2) 52-56

In order to meet the manufacturing process of micro part, where the space is small the production speed, temperature and vibration frequency are high, a novel method, which is used to measure the two-dimensional (2-D) sizes of the micro part fast and on-line, was proposed based on optical fibre image bundles and CCD camera. Double parallel lights were projected on the part symmetrically. The outline information of the part image on CCD camera through optical fibre image bundles and the part 2-D sizes can be obtained from the CCD camera. The optical fibre image bundles can be embedded into tool-set or fixed in limited space due to its small size, flexibility and bendability. Therefore, the method overcomes the bad influences from the part hard processing conditions effectively. Experiments

were carried out to measure workpiece with dimensions of 6.124 mm x 0.424 mm. The results show that the measurement time was less than 0.2 second and the accuracy was up to 25µm.

Keywords: On-line Measurement, Micro-part, Optical Fibre Image Bundles, Two-dimensional Sizes, CCD

Meng Xian-Yong, Liu Yang, Gao Yu-Xiao, Wang Mei, Liu Zhi-Hui Basic set of equations in mesoscale meteorological model associated with local fractional derivative operators involving the cantor and cantor-type spherical coordinates

Computer Modelling & New Technologies 2014 18(2) 57-62

Recently, the local fractional calculus theory was applied to process the non-differentiable phenomena in fractal domain. The main object of this paper is to present the basic set of equations in mesoscale meteorological model on the Cantor sets involving local fractional derivative operators and the corresponding cantor-type spherical coordinate equations. It is shown that these equations are efficient and accurate for describing some of atmospheric motion.

Keywords: Mathematic, basic set of equations, mesoscale meteorological model

Junsheng Jiao DSM control strategy of solid oxide fuel cell distributed generation system

Computer Modelling & New Technologies 2014 18(2) 63-69

The distributed power generation system of fuel cell has received the widespread attention in recent years for its pro-environment and high efficiency. The solid oxide fuel cell (SOFC) is a new power system, which turns chemical energy of fuels into electrical energy directly at middle and high temperature. Mathematical model of SOFC is analysed and the main circuit structure of independent power generation system is designed in this paper. The traditional PI control is adopted in DC/DC boost circuit to ensure the fuel cell can provide a stable DC output voltage. The three-phase inverter is used in DC/AC circuit and the dynamic equation derived from three-phase inverter is transformed to the synchronous rotating coordinate system. An innovative discrete sliding mode (DSM) control technology is applied and the system output stable sinusoidal AC voltage. The simulation experiments prove that SOFC power systems, in a certain range, can quickly and dynamically trace the change of the input voltage and load using DSM control.

Keywords: SOFC, Discrete Sliding Mode, Control, DC/AC, Converter

Jing Guoqing, Liu Guixian, Xu Yang, Zhang Jiong Micro-analysis of sub-ballast direct shearing under normal stress

Computer Modelling & New Technologies 2014 18(2) 70-75

This paper presents sub-ballast material analysis through Discrete Element Modelling (DEM) method to investigate its direct shear tests behaviour. Laboratory direct shear box tests were conducted on granite sub-ballast aggregate samples under different normal stress. The size, particle size distribution (PSD) properties of the sub-ballast particles were considered in the analysis. Simultaneously, direct shear box DEM simulations were conducted for different normal stress conditions, the sub-ballast micro-characteristics of the direct shearing tests were analysed, the contact force, displacement etc. index variation interrelated with the direct shearing tests conditions. The shear stress-shear displacement curves predicted from the DEM simulations were in reasonably good agreement with laboratory test results at all normal stress levels that shear box tests were conducted at. The results of discrete element modelling and tests of direct shear test for sub-ballast were presented. Such analysis may provide useful information on the understanding of sub-ballast with discrete element models. The micro-analysis and the direct shear tests were of importance to investigate the sub-ballast behaviour under the track system.

Keywords: direct shearing test; DEM; sub-ballast, normal stress

Liu Xiaoyong, Su Tiexiong, Zhang Yi Numerical analysis and improvement of torsional vibration of shaft systems for engine with cylinder deactivation

Computer Modelling & New Technologies 2014 18(2) 76-82

Cylinder deactivation is one of effective ways to improve fuel economy of engine, but will lead to changes in torsional vibration characteristics of shaft systems for engine. A lumped parameter model of torsional vibration of shaft systems for engine with cylinder deactivation was established, the numerical computing method was determined, harmonic analysis was engaged for the excitation torque of single cylinder. Based on these studies, torsional vibration of a V8-engine was analysed, the natural frequency results was verified by comparing with that of utilizing software AMESIM. The forced vibration results show that the torsional angle displacement of crankshaft under cylinder deactivation increases obviously, which mainly consists of the 2nd order rolling vibrations, but torsional stress decreases little. In order to control the rolling vibration, the measure of increasing the rotational inertia of the flywheel was adopted. The

results after the adjustment show that the vibration of crankshaft was under control. In a word, the method is feasible and referred.

Keywords: Cylinder Deactivation, Torsional Vibration, Natural Frequency, Forced Vibration, Rolling Vibration, Excitation Torque

Gong Yusheng Simulation design of bursa-wolf coordinate transformation model based on the access

Computer Modelling & New Technologies 2014 18(2) 83-86

With the popularity of GNSS technology, problems of conversion between different spatial rectangular coordinate systems are often encountered, which are difficult to solve. This paper introduces Bursa-Wolf coordinate transformation model in detail, this model is widely used in the conversion between different spatial rectangular coordinate system, In order to improve the efficiency of calculation, This paper writes a simulation program using tables, forms, macros and VBA in Access, also the simulation program has been verified by production instances, and the reliability of its results is ensured.

Keywords: Access, VBA, Spatial rectangular coordinate system, Bursa-Wolf Model, Simulation

Nahla Mohamed Ahmed, Ling Chen New approaches for link prediction in temporal social networks

Computer Modelling & New Technologies 2014 18(2) 87-94

Link prediction in social networks has attracted increasing attention from various domains such as sociology, anthropology, information science, and computer sciences. In this work, efficient approaches to predict potential link in temporal social networks are presented. One approach is based on reduced static graph using a modified reduced adjacency matrix to reflect the frequency of each link. Another approach is based on indices integration and exploits both the temporal and topological information. The approach integrates the indices in all the time steps, which reflect the topological information, and uses a damping factor to emphasize the importance of more recent links. Experimental results on real datasets show that our approaches can efficiently predict future links in temporal social networks, and can achieve higher quality results than traditional methods.

Keywords: Social networks, temporal networks, link prediction

Liu Ranran, Li Zhengming Modelling study of p-xylene oxidative side-reaction based on Bayesian filtering

Computer Modelling & New Technologies 2014 18(2) 95-98

The combustion loss model of HAC and PX was established based on Bayesian filtering. Moreover, this neural network model regarded the main adjustable process parameters (residence time, the concentration of Co, Mn, Br, reaction temperature and solvent ratio) as the independent variables and the total content of CO and CO₂ as the dependent variable. The simulation results show that the network is well performed. The effects of process operating parameters on HAC and PX combustion loss are analysed based on the model. Wherein, the effects of residence time, reaction temperature on the combustion loss are bigger, while the effects of other factors are relatively small.

Keywords: Mathematical model, PX oxidative side-reaction, Bayesian filtering, soft sensor

Huang Li Bipartite graph model for RDF data cleansing

Computer Modelling & New Technologies 2014 18(2) 99-106

Many systems use RDF to describe information resources and semantic associations between resources. RDF data plays a very important role in advanced information retrieval. Due to diversity and imprecise of resources, duplicates exist in RDF data. The query and retrieval of RDF data are studied by many researchers. However, researchers seldom study RDF data cleansing. In this paper, we focus on RDF data cleansing. According to the features of RDF data, we propose a new approach. This approach combines similarity and connections among resources. First, we introduce an intermediate model, named RDF-Bipartite Graph model, to represent the RDF data. This model improves from Bipartite Statement-Value Graphs model. Then on the proposed model, we design a Subgraph-Extend method, to find the path connecting two nodes. This method detects the minimum subgraph containing two nodes for connect-path finding. It avoids the connect-weight setting in traditional method. Experiments on publication datasets show that the proposed method is efficient in duplicate detection of RDF data, and has high performance and accuracy.

Keywords: RDF, data cleansing, Bipartite Graph

Information and Computer Technologies

Yang Xianhui, Ren Honge, Jing Weipeng A dynamic MAC protocol for wireless sensor network based network traffic monitoring and feedback

Computer Modelling & New Technologies 2014 18(2) 107-113

WSN application in coal mine gas monitoring protects safety production. In order to save energy consumption, SMAC use periodic listen and sleep mechanism, against the problem of fixed duty cycle causes increased delay, this paper puts forward a new MAC protocol—DSMAC(Dynamic SMAC)which can feedback the channel's congestion level reasonably and adjust the duty cycle dynamically. The dominant idea of the new scheme is to record the packet amounts in the sending stage by sensor nodes, and according to the packet sending rate every sensor node can change its listening time adaptively when synchronization cycle finished to minimize delay and collisions, saving energy and channel resources. The result of simulation shows that compared with other MAC protocols, the DSMAC protocol can improve network throughput and reduce energy consumption in coal mine gas monitoring environment.

Keywords: WSN, periodic listen and sleep, duty cycle, DSMAC

Guoyan Yang, Xin Guan A concurrent MAC protocol based on location information in wireless sensor networks

Computer Modelling & New Technologies 2014 18(2) 114-119

The traditional CSMA MAC node simply blocks its transmission if the medium is sensed to be busy. Thus, it is inefficient in terms of the network throughput due to overcautious estimation of the interference. In this paper, we propose a novel location-aware medium access protocol for data intensive wireless sensor networks. In this protocol, the contending nodes make use of their location information to achieve the concurrent transmission of exposed terminal so as to reduce collisions and improve the overall performance. We evaluate it in terms of delay and throughput and compare it with S-AMC using simulations. Results show that the proposed MAC protocol can take advantage of the location distribution of nodes to improve the average throughput of the network, reducing data transmission delay, and effectively improving the efficiency and performance in data intensive wireless sensor networks compared to S-AMC.

Keywords: exposed terminal, MAC protocol, wireless sensor networks, concurrent transmission

Huisong Li Flow-insensitive type qualifier inference on programming languages allowing type casts

Computer Modelling & New Technologies 2014 18(2) 120-125

Type qualifiers are lightweight specifications of atomic properties that refine the standard types. Flow-insensitive type qualifier inference has been used in the CQual framework to improve the quality of C programs. However, type casts will affect the effectiveness of type qualifier inference, as they can lead to either accepting some flawed programs due to discarding some useful qualifier information, or rejecting some safe programs when the analysis is conservative. In this paper, we first present a language, which allows type casts and formalize its flow-insensitive qualifier inference system. We then show some examples to illustrate how qualifiers are lost because of type casts in CQual and give an idea on solving this problem.

Keywords: Type cast, Type inference, Flow-insensitive type

Lichuan Gu, Jianxiao Liu, Chengji Wang Fragments based tracking with adaptive multi-cue integration

Computer Modelling & New Technologies 2014 18(2) 126-134

In this paper, we address the issue of part-based tracking by proposing a new fragments-based tracker with multi-cue integration. First the target template is divided into multiple fragments to get the contribution of each fragment to the possible positions of the target in the current frame, similarity measure is used in edge histogram and HSV histogram in every fragment, and the weights of cues integration are computed adaptively. Second we present a fragment template update mechanism with the discrimination between occlusions and appearance changes. The template is unchanged when the target is occluded and some fragments of template are updated in the case of appearance changes. In the experiments we use the indoor and outdoor test videos which contain the illumination changes, occlusions, and the appearance changes of targets. The experimental results show that our approach has strong robustness and high tracking accuracy.

Keywords: Target tracking, Multi-cue integration, Fragments tracking, Template update

Shaobo Shi, Qi Yue, Qin Wang FPGA based accelerator for parallel DBSCAN algorithm*Computer Modelling & New Technologies 2014 18(2) 135-142*

Data mining is playing a vital role in various application fields. One important issue in data mining is clustering, which is a process of grouping data with high similarity. Density-based clustering is an effective method that can find clusters in arbitrary shapes in feature space, and DBSCAN (Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise) is a basic one. With the tremendous increase of data sizes, the processing time taken by clustering algorithms can be several hours or more. In recent years, FPGA has provided a notable accelerating performance in data mining applications. In this paper, we study parallel DBSCAN algorithm and map it to FPGA based on the task-level and data-level parallelism architecture. Experimental results show that this accelerator can provide up to 86x speedup over a software implementation on general-purpose processor and 2.9x over a software implementation on graphic processor.

Keywords: Data mining, Clustering, Parallel DBSCAN, FPGA, Hardware Accelerator

Kanghong Duan, Hongxin Zhang, Shilin Song, Peigang Wang Micro real-time pre-emption operating system for industry wireless sensor networks*Computer Modelling & New Technologies 2014 18(2) 143-150*

Event-driven systems and thread-driven systems are two major design philosophy of operating system in wireless sensor networks. Systems based on multi-threaded are more timeliness than the event-driven systems, which can meet the requirements of time-critical tasks by means of task pre-emption, while systems based on event-driven are more energy efficient. Furthermore, μ COS-II is a classical system, which combines benefits in both systems. Therefore, our recent work we have shown that a micro real-time pre-emption operating system has been proposed on the basis of μ COS-II. First of all, a clear hardware abstraction layer (HAL) is given to combine the kernel and hardware in the system architecture. Moreover, this system is more capable of fitting both sensor network design goals of energy efficiency and timeliness. We are dedicated to modify the existing system from the scheduling strategy and data structure aspects, which lead to the performance of the modified system largely improved. Above all, the performance of our operating system is better than the original μ COS-II and TinyOS from task switch time, FLASH usage and RAM usage perspectives.

Keywords: Wireless Sensor Network, Operating System, Pre-emption, Improvement

Ao Jianfeng Ground crack extraction in mining subsidence areas based on point cloud data*Computer Modelling & New Technologies 2014 18(2) 151-156*

During the underground mining of mineral resources, some ground points within a certain range above the working face will migrate correspondingly. The inconsistent horizontal migration of various ground points, due to different impacting time and degree, will lead to deformation of the ground surface, and when it reaches a certain degree, discontinuous damages to the continuous ground surface will occur, which are shown as cracks on the ground surface. In this paper, to extract planar cracks contained in the point cloud data, the point cloud data was firstly projected in the elevation direction for 2D processing and then a thickening algorithm for scattered point clouds was proposed. The point cloud data at different distances from the instrument was thickened block by block by choosing different thickening windows in order to highlight the crack data in the 2D image and to obtain the point cloud image applicable to the regular edge detection algorithm. Finally, the Canny operator was used to extract the edge information of cracks. In the end, the algorithm was tested in engineering projects and proved to be highly effective in extracting crack data.

Keywords: Crack Extraction, Point Cloud, Mining, Subsidence

Lou Ning, Zheng Xiaobo, Yang Yongchong Expression model of multi-resolution 3D geographical space*Computer Modelling & New Technologies 2014 18(2) 157-160*

Aiming at improving deficiencies in multi-resolution two-dimensional expression of geographic information method, this paper studied method of multi-resolution 3D expression of geographical information, putting forward grid method to express different regional landforms separately and a method of expressing different regional ground objects on level of detail in three aspects as shape, texture and properties, plus an example to simulate and illustrate. Multi-resolution 3D expression of geo-information has broad application prospects in smart city and city planning, which has a certain exploring value for the massive information, for its expression and visualization.

Keywords: Multi-resolution, 3D visualization, Geographical Space Model, 3D terrain model, 3D ground object model

Chen Qing, Yong Zhong, Liuming Xiang Cloud database dynamic route scheduling based on polymorphic ant colony optimization algorithm

Computer Modelling & New Technologies 2014 18(2) 161-165

Big Data era spawned the development of Cloud database. As a database, which need easily scale out, how to quickly find the available nodes are focuses of the study. Ant colony algorithm is based on bionic optimization algorithm and has the characters of smart searching, global optimization, robustness, distributed computing and easily combined with other algorithms, but the algorithm is prone to premature convergence, making the results often caught local optimum. According to this, polymorphic ant colony algorithm was proposed which combined with a Cloud database; the algorithm can quickly and reasonably find the nodes in Cloud environment, reducing the load of routing, thus greatly improved the Cloud database's ability of scaling out.

Keywords: Big Data, Cloud database, Ant colony algorithm, premature convergence, scaling out

A Aitmagambetov, Yu Butuzov, S Torekhan Matters of satellite queuing network design in K_a -band for Republic of Kazakhstan

Computer Modelling & New Technologies 2014 18(2) 166-169

This work was carried out within the framework of research opportunities of using K_a - band satellite communication systems in the Republic of Kazakhstan. The paper deals with the multi-beam coverage in Kazakhstan (the distribution of beams in area and determine their capacity), as well as evaluation of the main parameters of subscriber channels. The need for this research was due to the fact, that the design of multibeam network for Kazakhstan is important to consider a distinct uneven distribution of the population (or potential users), low average density (about 6 persons / sq km.) and a fairly significant differences rain intensity in some areas of the territory.

Keywords: satellite networks, K_a -Band, multi-beam antennas, bandwidth, signal-code constructions

Operation research and decision making

R Uskenbayeva, B Kurmangaliyeva, N Mukazhanov Development of multidimensional model of data for information and analytical decision-making support system

Computer Modelling & New Technologies 2014 18(2) 170-174

Today almost in all spheres of industry it is observed the urgent need of information technologies development and increase of using opportunities. This paper is devoted to creation of multidimensional model of data for information and analytical system of decision-making support for management. The information and analytical system is developed by means of OLAP technology (On-Line Analytical Processing) which in turn analyses collected materials at a real time based on strategic level of solution. In information systems this technology displays data in the form of a multidimensional cube, and necessary data for users will be expressed in the form of cube slices. The multidimensional model of data is developed for information system intended for decision-making support of higher education institutions management.

Keywords: Multidimensional model of data, data warehouse, hypercube, dimension, higher education

Shangyao Shi, Shiping Zhao, Jun Li On tracking ability analysis of linear extended state observer for uncertain system

Computer Modelling & New Technologies 2014 18(2) 175-179

It is known that the linear active disturbance rejection control (LADRC) is very an effective approach to control the uncertain systems. The linear extended state observer (LESO) is the major link of the LADRC, so this paper presents a modified LESO, which is used to track the state variables and estimate the unknown total disturbance. Furthermore, this paper redefines the "time-scaling" of the plant, which is a function with respect to the amplitude of the unknown total disturbance. It is first time to present the specified formula of the maximum sampling-period of LESO for some existing plants. On the another hand, the tracking ability of the designed LESO is quantitatively described in this paper. The discussions and analysis, especially the quantitative formulas presented in this paper, will help the scholars and engineers to design the LESO in practice.

Keywords: LADRC, LESO, Modified LESO, "Time-Scaling", Sampling-Period

Wei Shang, Maozhu Jin, Jie Wang Using genetic algorithms on service facilities layouts design*Computer Modelling & New Technologies 2014 18(2) 180-185*

Traditionally, the objective of a service facility layout problem is to minimize the material handling cost and optimize the physical spaces that support a productive activity of the service system. Service delivering interface has direct impact to the efficiency and effectiveness of service system in light of customer's contact. Layout problem of service facilities has received little attention in the literature. Due to complexity and combination of the facility layout problems, the genetic algorithm is put forward that codifies a chromosome and uses a direct representation to assign the departments to the different branch areas, enables good solutions to be justified in real instances of the problem in tourist service centre. The solution is shown on a decision plan representing the graphic layout of the departments within the service facility. This application is highly flexible that the decision maker can accordingly define the departments' number to be accommodated, facility sizes, segment sizes, and algorithm parameters to be added or simplified.

Keywords: Layout design, Service facility, Genetic algorithm, Tourist service centre

Wu Shimei, Pei Yulong, Cheng Guozhu Method of traffic zone division based on spectral graph theory*Computer Modelling & New Technologies 2014 18(2) 186-191*

Aiming at the macro planning of traffic district division, the spectral graph theory, graph, spectral and matrix (i.e. Laplace matrix) were introduced, and the method and procedures of the traffic zone division based on spectral graph theory were put forward. Euclidean distance between communities was calculated according to four indexes, i.e., economic indicators, land development intensity, residential population quantity and spatial distance. The balance degree of indexes between communities was regarded as criterion to evaluate community division precision. Taking Dongguan, a city in the south China, as an example, above mentioned method was applied, conducting the traffic zone division, with rail transit network planning flow prediction.

Keywords: Traffic zone division, Spectral graph theory, Euclidean distance, Dongguan

Yanghe Feng, Baoxin Xiu, Zhong Liu A dynamic optimization model on decision-makers and decision-layers structure (DODDS) in C2-organization*Computer Modelling & New Technologies 2014 18(2) 192-198*

The highly-complexity, environmental uncertainty, and structure changes bring more requirements for the agility and resilience of its core command and control C2-organizations. In order to better understand such organization's dynamic and emergence behaviours as a system of systems, we establish time-domain based metric model to evaluate C2 organizational decision-making capability. We develop an optimization model of organizational structure. The model is based on decision-makers and decision layer dynamics. The model aims at helping gain an optimal organizational structure with higher operational flexibility, low cost and high performance.

Keywords: Operational System of Systems (SoS), Command and Control, Adaptive Optimization, Simulated Annealing

Zhao Suxia, Yang Xuemei, Sun Peihong Research on the method of determining the service scope of civil airport – the case of Chengdu Shuangliu airport*Computer Modelling & New Technologies 2014 18(2) 199-204*

Airport hinterland is the most important of influencing factors for civil aviation airport layout. By using methods of breaking point, expert consultation and entropy weights, the formula of computing the acting radius for airport hinterland was set up according to breaking point theory, and the idea and method were proposed to determine the service scope of an airport in an economic region. Then demonstrate the rationality of airport layout in the Chengdu-Chongqing economic zone by calculating the actual case of airport in the Chengdu-Chongqing economic zone. And get a conclusion that Leshan airport is inappropriate expansion in the near future.

Keywords: Civil airport, Breaking point theory, Entropy weights, Service scope

Liu Zhusheng Tourist flow-control study based on utility*Computer Modelling & New Technologies 2014 18(2) 205-211*

The key to solving tourist congestion during peak seasons is to improve the level of management and control the tourist flow in a scientific manner. Scenic capacity is usually set to a fixed value. Consequently, many plans have been developed, but have often produced poor results in practice. The utility concept in economics was introduced in this

study, according to the law of diminishing marginal utility; then a scenic spot utility function was proposed. With the aim of maximizing the utility value, a mathematical programming model was developed. Through the process of solving the example problem, it was shown that, in a certain number of tourists, scenic managers can optimize each visitor route to upgrade the tourist experience by effectively managing the flow of visitors.

Keywords: utility, tourist flow control, scenic management, tourism modelling

Liu Caihong, Xiong Wei The agent – based warning modelling of internal quality risks in supply chain for manufacturer

Computer Modelling & New Technologies 2014 18(2) 212-220

With the dynamic of market and deepening competing, enterprises will to be faced with much uncertain quality risks in supply chain. At present, the quality risk has been seen as one of the most important risks of supply chain, which is the most difficult to prevent and manage. To do it, the entities-attributes model based on the business of manufactures was used as the reference data sources of building evaluation index system for internal quality risks of supply chain. Because the early warning of risks manually has the limits of information insensitivity and risks identification slowly, this is apt to cause the delay of risks precautions. Therefore, the intelligent Agent based modelling method will be applied to construct the four early warning situations based warning model according to the internal risk of supply and demand for manufacturers. In addition, the evaluation algorithm, the rules of early warning and the running process for this Agent were focused in the study. This study will play a reference role to the analysis and management on the quality risks of supply chain.

Keywords: Quality risks, Supply chain (abbr. SC), Agent, Manufacturing enterprise, Early-warning modelling

Cong Guodong A study on the lock-in risk in IT outsourcing projects: the mechanism and the control system

Computer Modelling & New Technologies 2014 18(2) 221-226

This paper, proposes an insight into the mechanism of lock-in risk, which is the primary risk and the greatest concern in managing IT outsourcing projects. Based on Transaction Cost Theory (TCT), it develops an integrated three-layer model to instruct the mechanism of lock-in risk, namely, what will lead to lock-in risk (risk drivers), and what lock-in risk will result in (risk consequences). Four risk drivers such as asset specificity and three risk consequences such as cost escalation are identified and discussed in detail. Meanwhile, it instructs why four kinds of control could mitigate lock-in risk both from the sources and the consequences. Therefore, the effectiveness and reliability is enhanced throughout the risk management process of IT outsourcing projects, particularly for dynamic risk identification, controlling and monitoring.

Keywords: project management, lock-in risk, IT outsourcing, risk management

Bingjie Li A game theory analysis - dumping by multinational company and antidumping in China

Computer Modelling & New Technologies 2014 18(2) 227-231

Based on a dynamic game model, this paper analyses dumping by multinational firm and Chinese government's antidumping behaviour. It is shown that no matter how many products are dumped in domestic market; Chinese authority should impose punitive damages against the foreign firm as long as the scale of antidumping duty is not too high to stop multinational company's investment. This strategy will improve social welfare.

Keywords: Multinational company, Dumping, Antidumping, Game

Zhihong Ma, Jianping Chen, Shouzhen Zeng A new method based on induced aggregation operator and distance measures for fuzzy decision-making

Computer Modelling & New Technologies 2014 18(2) 232-239

In this paper we introduce a new fuzzy decision making model that unifies induced aggregation operators and distance measures in the same formulation. We develop the fuzzy induced generalized ordered weighted averaging distance (FIGOWAD) operator. The main advantage of this operator is that it provides a parameterized family of aggregation operators between the minimum and the maximum and a wide range of special cases. Another advantage is that it is able to deal with the fuzzy environment where the information is very imprecise and can be assessed with interval numbers. Moreover, it uses induced aggregation operators that provide a more general representation of the attitudinal character of the decision-maker. We study some families of the FIGOWAD operators. We end the paper with an

application of the new approach in a business decision-making problem about the selection of strategies.

Keywords: Fuzzy numbers, induced aggregation operator, distance measure, decision making

Bi Jianxin, Lei Lianghai Analysis on the effectiveness of China's macro-economic policy based on the modified mundell-fleming model during the post-financial crisis period

Computer Modelling & New Technologies 2014 18(2) 240-249

This paper proposed the hypotheses of Mundell-Fleming model applicable to current Chinese economic environment, modified the traditional Mundell-Fleming model and analysed the effectiveness of fiscal and monetary policies under different exchange rate systems. Under a fixed exchange rate system, the monetary policy causes economic instability and aggravates economic inequality, thus increasing the difficulty in policy intervention, in which case the fiscal policy has significant results only with the positive coordination of monetary policy; under a floating exchange rate system, the fiscal policy has remarkable effects and the monetary policy has effects which are not uncertain. Finally, the paper analyses the effectiveness of China's macroeconomic policy using the modified Mundell-Fleming model and proposes the orientation for China's macroeconomic policy in post-crisis period.

Keywords: Mundell-Fleming Model, Sub-prime Mortgage Crisis, Financial Crisis, Fiscal Policy, Monetary Policy

Sheng Yao, Longshu Li Approximation reductions in an incomplete variable precision multigranulation rough set

Computer Modelling & New Technologies 2014 18(2) 250-258

This paper deals with approaches to the granular space reductions in the variable precision multigranulation rough set model. The main objective of this study is to extend four kinds of the granular space reductions called a tolerance relations optimistic multigranulation β lower approximation distribution reduction, a tolerance relations optimistic multigranulation β upper approximation distribution reduction, a tolerance relations pessimistic multigranulation β lower approximation distribution reduction and a tolerance relations pessimistic multigranulation β upper approximation distribution reduction, which preserve the optimistic/pessimistic multigranulation β lower/upper approximation distribution of the decision classes. Some judgement theorems are investigated. The example proves that the new variable precision multigranulation rough set model can effectively deal with incomplete information, from which we can obtain approaches to the granular space reductions of incomplete decision systems in variable precision multigranulation rough theory.

Keywords: Approximation Reduction, Tolerance Relation, Variable Precision Rough Set, Multigranulation

Wang Hongbao Cost-sensitive back-propagation neural network for financial distress prediction

Computer Modelling & New Technologies 2014 18(2) 259-264

Financial distress prediction (FDP) models, which classify financially distressed companies from healthy ones, prevent market participants from suffering economic loss. In the process of FDP, the misclassification of type I error of the model incurs much higher cost than that of type II error. Most of the previous FPD models do not take the asymmetric costs into consideration. In this paper, cost-sensitive back-propagation neural network (CS-BPNN) FDP model is proposed for minimizing the cost of prediction error such that the loss of users of the model will suffer less. The performance of the model is evaluated by taking 180 Chinese listed companies as sample data and adopting 8 times of sampling to assess different misclassification costs and prediction accuracy. The experimental results suggest that the proposed approach helps to improve the prediction performance in asymmetric cost setup.

Keywords: Financial Distress Prediction (FDP), Cost Sensitive Learning, Back-propagation Neural Network (BPNN), Cost of Prediction Error, Cost-sensitive Back-propagation Neural Network (CS-BPNN)

Shaoren Wang, Zujun Ma, Bochao Zhuang Fuzzy location-routing problem for emergency logistics systems

Computer Modelling & New Technologies 2014 18(2) 265-273

To optimize the post-earthquake emergency logistics system with two-echelon multi-facilities, this study develops a model for fuzzy location-routing problem by considering fuzzy demand of relief materials, timeliness and limited resources. The goal of the model is to minimize the total cost and the relief time of system. Furthermore, this research proposes an improved genetic algorithm based on weighted coefficient transformation. The result of a numerical example shows that the model and algorithm are effective for resolving the joint decision-making of facility location-

allocation problem and vehicle routing problem in post-earthquake.

Keywords: Emergency Logistics, Location-routing Problem, Earthquake Disasters, Multi-objective Optimization, Improved Genetic Algorithm

Peizhe Li, Lirong Jian, Kun Zhang, Shanshan Pei Evaluation of regional economic development level based on grey clustering and rough set

Computer Modelling & New Technologies 2014 18(2) 274-280

The evaluation result of regional economic development is an important basis for government to make regional economic development planning, scientific regional economic development evaluation system and method has an important significance for understanding the regional economic development. Using grey clustering to classify different regions by the new evaluation index system, and using the rough set theory to derive fuzzy decision rules, these rules can explain the preference behaviour of decision makers and provide scientific and rational decision-making suggestions.

Keywords: Regional economic, Grey fixed weight clustering, Dominance rough set, Evaluation, Decision rules

Zhan'an Zhang, Xingguo Cai Power purchase plan using minimal cost flow

Computer Modelling & New Technologies 2014 18(2) 281-285

The minimal cost flow method presented in this paper is used to calculate the minimum power purchase cost, because the user's goal is the minimal power purchasing cost under open electricity market. By establishing the minimal power, purchasing cost model and using the minimum cost flow algorithm to solve the objective function, the method is used to calculate a network in a single time. A continuous period of power purchase plan is obtained by accumulating each period network flow together. Example analysis proves that it is feasible to solve the problem with the minimal cost flow algorithm.

Keywords: Power Purchase Cost, Minimal Cost Flow, Network, Electric Market, Power Purchase Plan

Innovative Education

Tatyana Koycheva Novel approaches to form the corporate culture in pedagogical universities

Computer Modelling & New Technologies 2014 18(2) 286-289

The role of a corporate culture in the development of the modern education, especially in the improvement of the university education is considered. On the example of pedagogical university it's shown that the active research work promotes the formation of a corporate culture within the educational institution. At the same time an availability of the corporate culture leads to the expansion and intensification of scientific work. As a result of our research it was confirmed a particular importance of scientific schools and international consortiums in the formation of a corporate culture.

Keywords: corporate culture, pedagogical universities, research work, scientific schools