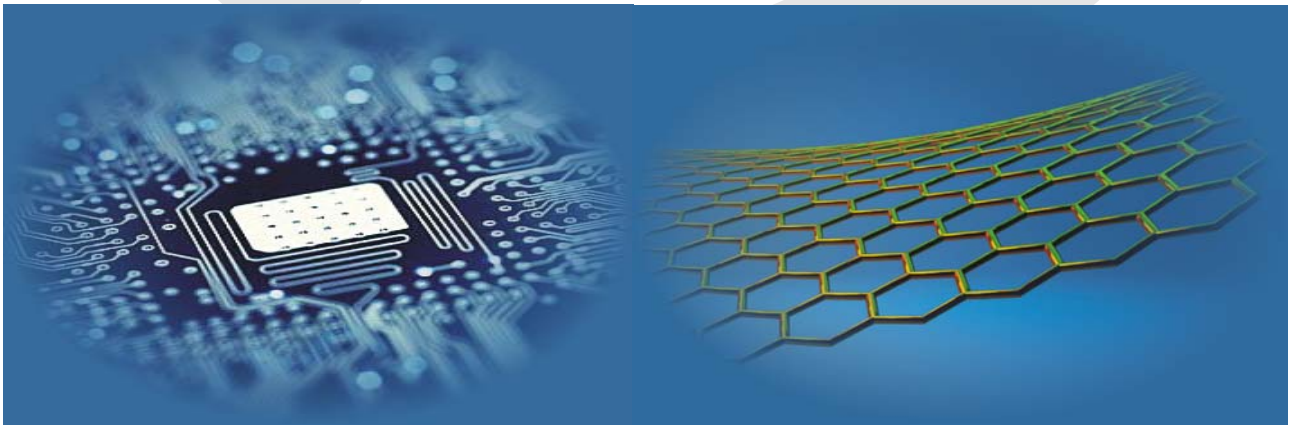
Transport and Telecommunication Institute
and
Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2014, Volume 18, No 1

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



Riga – 2014

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

Transport and Telecommunication Institute

1 Lomonosova, **Bld 4,** LV-1019, Riga, **Latvia**
Phone: (+371) 67100593
Fax: (+371) 67100535
E-mail: cm&nt@tsi.lv
www.tsi.lv

COMPUTER MODELLING AND NEW TECHNOLOGIES, 2014, Vol. 18, No.1
ISSN 1407-5806, ISSN 1407-5814 (on-line: www.tsi.lv)

Scientific and research journal
The journal is being published since 1996

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/
EI Compendex

Content

Editors' Remarks

****

All the World's a Stage

by William Shakespeare

All the world's a stage,
And all the men and women merely players;
They have their exits and their entrances,
And one man in his time plays many parts,
His acts being seven ages. At first, the infant,
Mewling and puking in the nurse's arms.

Then the whining schoolboy, with his satchel
And shining morning face, creeping like snail
Unwillingly to school. And then the lover,
Sighing like furnace, with a woeful ballad
Made to his mistress' eyebrow. Then a soldier,
Full of strange oaths and bearded like the pard,

Jealous in honor, sudden and quick in quarrel,
Seeking the bubble reputation
Even in the cannon's mouth. And then the justice,
In fair round belly with good capon lined,
With eyes severe and beard of formal cut,
Full of wise saws and modern instances;

And so he plays his part. The sixth age shifts
Into the lean and slippered pantaloon,
With spectacles on nose and pouch on side;
His youthful hose, well saved, a world too wide
For his shrunk shank, and his big manly voice,

Turning again toward childish treble, pipes
And whistles in his sound. Last scene of all,
That ends this strange eventful history,
Is second childishness and mere oblivion,
Sans teeth, sans eyes, sans taste, sans everything.

William Shakespeare (**1564-1616**) [♣]

****

[♣] **William Shakespeare,** 26 April 1564 - 23 April 1616 / Warwickshire, an English poet and playwright, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. He is often called England's national poet and the "Bard of Avon". His surviving works, including some collaborations, consist of about 38 plays, 154 sonnets, two long narrative poems, and several other poems. His plays have been translated into every major living language and are performed more often than those of any other playwright.

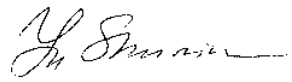
Modern technology requires educated work force and hence imperative for educated population. The needs of new emerging technologies and a beneficial state of society are compatible in this case. There is no monolithic thing called technology. Rather there are various technologies, which converge or compete to fit into what can be called an ecosystem of technological and societal arrangements. Societal and technological arrangements co-evolve. This co-evolution happens most favourably in an educated, intellectual, and affluent society that is tolerant of change and divergent views. By fostering an educated, intellectual society, it creates conditions that foster responsible moral and social behaviour of the individual and contributes to shaping intellectual humankind.

The map of the globe does not show a place called Technoworld, yet in many ways we are already its citizens and consumers. Moreover, as consumers, our lives and behaviours are strongly influenced and shaped by an ever-increasing system of modern technology, transcending national boundaries to create a vast technological synergy grounded in complex applications in industrial production, logistics, electronic communications, agribusiness, medicine, and science. We begin to realise that, like it or not, we have become a part of technoculture. The industrialized world has developed into a production and consumption global community with a highly advanced level of technologies. This process has led to markedly increased demands for a standard of living and consumption.

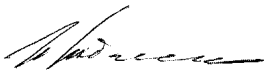
This 18th volume No.1 presents actual papers on main topics of Journal specialization, namely, Operation Research and Decision Making, Information and Computer Technologies, Mathematical and Computer Modelling, Nanoscience and Nanotechnologies.

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

Generalized model of pulsating track device

N. Mykytenko[1], D. Fink[2, 3], A. Kiv[1, 2*]

[1]*South-Ukrainian national pedagogical university after K. D. Ushinskij, 65020 Odessa, Ukraine*

[2]*Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva, 84105 Israel*

[3]*Division de Ciencias Naturales e Ingeneria, Universidad Autónoma Metropolitana-Cuajimalpa, Artificios 40, Col. Hidalgo, Del. Álvaro Obregón, C.P.01120, México, D.F.; México*

Received 10 January 2014, www.tsi.lv

Abstract

A model for description and study of pulsating track-based devices is developed. The track electronics opens up perspectives for solving new scientific and technical problems. The successful solution of these problems requires an elucidation of the mechanisms of the functioning of track-based devices. In this paper, the nature of the pulsating behaviour of electric parameters of track devices is clarified using a specially developed model based on classical Molecular Dynamics. It is demonstrated that the model describes adequately the main features of pulsations in track devices that were established experimentally.

Keywords: track electronics, current pulsations, molecular dynamics

1. Introduction

Nowadays nanotechnologies develop into various directions and nanostructures have numerous application scopes. Among the new areas of nanoelectronics, which have recently been widely developed, there is also the ion track electronics. It refers to new directions with promising perspectives and points at the emergence of electronic devices with unique properties not previously known in traditional electronics. This may open new possibilities of solving problems in the electronics industry [1-4]. In particular, track devices are used for creation of novel biological sensors, which can identify many biological objects [5-8]. Figure 1 shows the principle setup to the corresponding experimental arrangements.

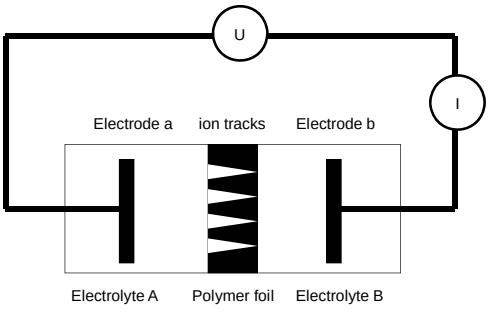


FIGURE 1. Principle arrangement of experimental setup to study current spike emission from ion track-containing foils embedded in electrolytes via current/voltage measurements

Track devices are complex systems with a set of parameters that depend on the sizes and shapes of tracks, their density, electronic structure of the inner surface of tracks, on the features of the film surface and so on. In experimental works of both S. Ziwy et al. and D. Fink et al. the effect of current spikes was found and studied for ions passing through electrolyte-filled individual etched [9] and multiple thin latent [10] and etched [11] ion tracks in polymers and other materials [12]. Later this effect was studied in a much more regular and controlled way than ever before [13]. In [14] a model of the pulsating track devices was proposed based on ideas of the theory of neural networks.

The operation of the track-device is determined by many microscopic processes, the simultaneous description of which is a difficult task. Therefore, a phenomenological description of such devices is an important stage in the clarification of the main principles of their functioning. In this paper, we report the results of creation and study of generalized phenomenological model of pulsating track device.

2. The model

A physical model not necessarily has a superficial resemblance to a real object, which has to be studied. A model should reflect the basic properties of the real object and the features of its behaviour. In the present paper, we envisage to investigate the pulsating behaviour of electrical parameters of track-based devices. In the case of pulsating track, devices the following main features are observed [13, 14]:

* Corresponding author: E-mail- kiv@bgu.ac.il

- Spikes emitted from latent or funnel-type tracks in polymers embedded in suitable electrolytes often show another peculiarity, which was hitherto unknown. Between both track types, a principle difference as concerns the spike emission does not exist.
- The spike emission depends on the amplitude and frequency of the applied voltage.
- The maximum spike heights do not seem to be affected markedly by the frequency of the applied voltage.
- The high ion track density is necessary to obtain the effect of spikes. This is caused by some specific interaction between tracks, mechanism of which has to be clarified.
- The spikes preferentially occur at pronounced, rather equidistant voltages.
- The spike spectra are not always reproducible though their principle features remain the same.
- With frequency decrease spike emission appears to vanish rapidly which indicates the existence of a threshold frequency for spike emission.

To develop a model we used a classical Molecular Dynamics (MD) with a Verlet algorithm [15, 16]. The model is a 3D space lattice, each node of which corresponds to an individual track in the track device, Figure 2.

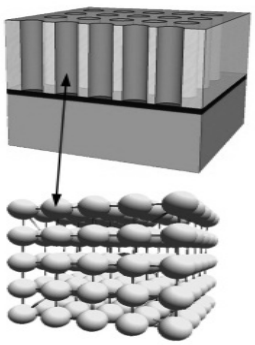


FIGURE 2. Correspondence between individual track and MP

The density of nodes is proportional to the density of tracks. Each node is a potential well, in which one or more so called model particles (MP) can be placed. Thus, in fact the model resembles a regular crystal lattice. Because the geometry of the model lattice is irrelevant, we use a simple cubic geometry. The lattice has a temperature, which is introduced in accordance with a conventional MD method [17]. The particles oscillate in the lattice nodes according to the introduced temperature. We suggest that the current value in the track device is proportional to the average amplitude of the oscillations of MP. The application of a sinusoidal voltage is modelled by the action of external forces (EF) introduced in the model, which at certain time intervals "nudge" the particles in the nodes. The directions of the action of EF are determined by the random function. We can vary the value and the frequency of the EF action, thus simulating the change of the applied voltage. The interaction of particles in the nodes is described by a potential of the most general form (Lenard-Jones) [18, 19].

The potential parameters, the particles mass and temperature are chosen so as to make the model stable and allow carrying out a computer experiment.

We performed a computer experiment to check how the developed model reflects the maim properties of pulsations in the track device. In computer experiment, it has been seen that with the increase of the absolute value of EF the average amplitude of MP oscillations increases. However, from time to time the mean amplitude of MP oscillations increases dramatically, Figure 3. This situation corresponds to the occurrence of the current spike in the track device, Figure 4. Visualization of the situation in Figures 3a,b shows that the observed MP oscillations spikes correspond to the cases when sufficient amounts of MP move simultaneously for a large distance from their nodes as a result of their interaction and return back after a short time. The conditions of such “model spike” are determined by the parameters of the interaction potential between MP, the action of EF and the temperature of MP. The temperature of the MP describes the fluctuating nature of the excited states of the MP. Below in Table 1 the correspondence between the model characteristics and the characteristics of the real track devices is shown.

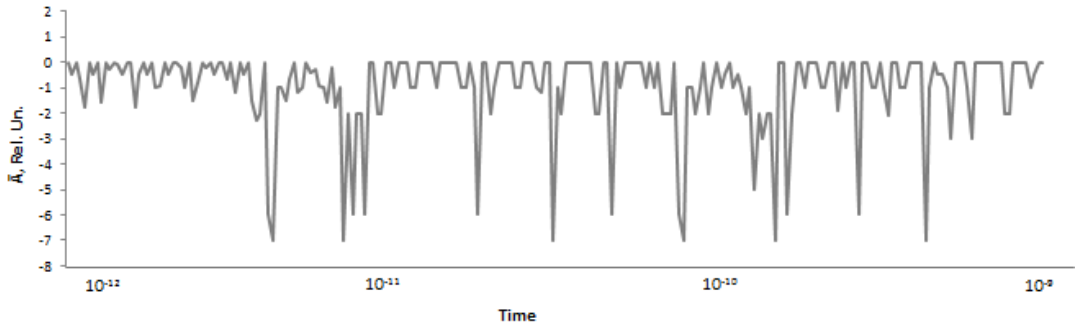


FIGURE 3a. Illustration of model spikes in the model experiment.
At the vertical axis is the average amplitude of MP oscillations. At the horizontal axis is the computer time in seconds

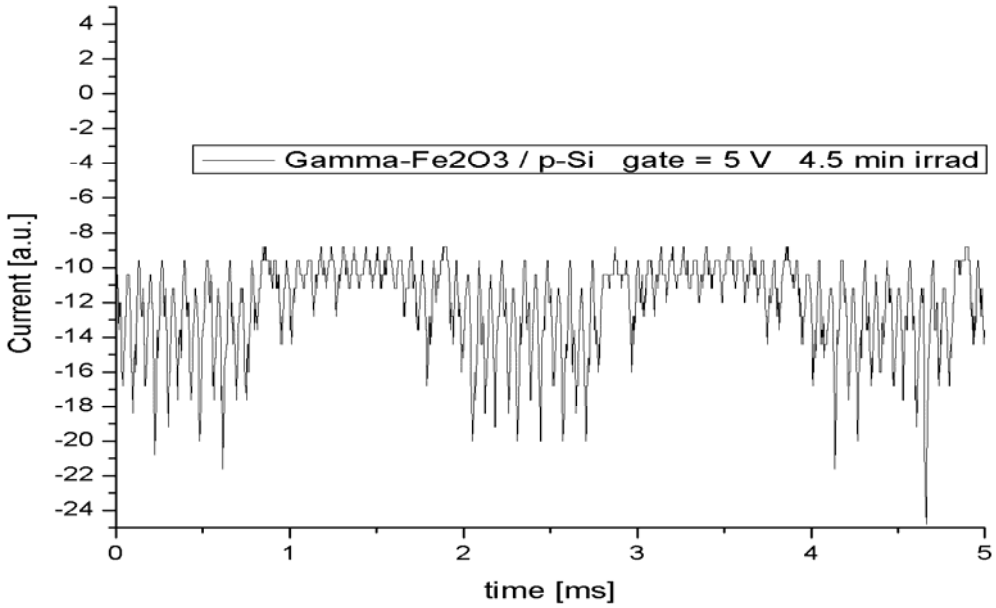


FIGURE 3b. Current spikes in conditions of real experiment with track device [12]

TABLE 1. Correspondence between the model characteristics and the characteristics of the real track device

Device characteristics	Model characteristics
Average spike height	Average amplitude of MP oscillations
Value of applied voltage	Value of EF
Frequency of applied voltage	Frequency of EF action
Time interval between current spikes	Time interval between MP oscillations spikes
Areal density of tracks	Spatial density of nodes

To realize the model we have implemented an application on C# .Net [20, 21] using abilities of powerful graphical engine Unity3d [22]. The program has a graphic interface that contains in the screen modelling lattice elements, two windows with program parameters and the information about the selected particles.

The program has two basic classes. The first one is MD that performs all calculations with MP located in the nodes and interacting by Lenard-Jones potential, and the second one is the Graphics that is responsible for rendering the results in real time. As the output data, MD program calculates forces, velocities and coordinates of MP. By scaling the data, we moved to higher order values of calculated quantities. The dimensionless equations were used with parameters of the appropriate order. An optimization algorithm for reducing the number of operators in the program was an important factor that reduced the accumulation of computational errors [4].

3. Results and discussion

Creating the model, we proceeded from the fact that in the cases of different forms of tracks (latent or funnel-type tracks) and different materials of films the main features of current spikes are the same. Thus, the idea of creating a phenomenological generalized model arose. It was necessary to create a model that would reflect the main features of the real track device. Computer experiment with the developed model led to the results shown in Figures 3a, 4a, 5a, 6 and 7. The corresponding experimental results are presented in Figure 3b, 4b and 5b.

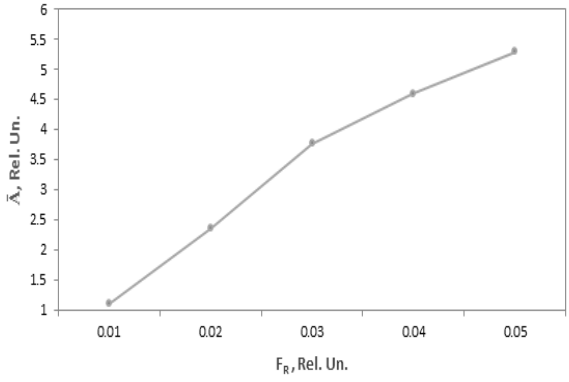


FIGURE 4a. Dependence of the average amplitude, $\bar{A}$ of MP oscillations on the value of EF

In Figure 3a we see that the model provides a good qualitative picture of current peaks observed in the real

track device (Figure 3b). The dependence of the average amplitude $\bar{A}$ of MP oscillations on the value of EF is displayed in Figure 4a. This dependence is in good agreement with the dependence of the average spike height on the value of applied voltage (Figure 4b).

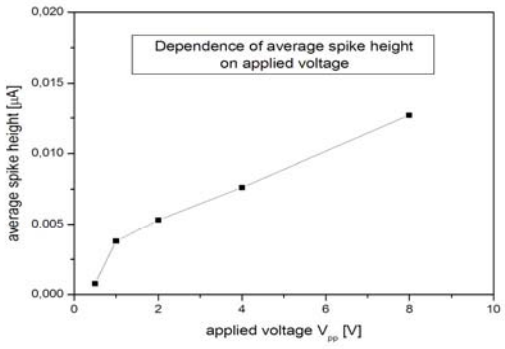


FIGURE 4b. Dependence of the average spike height on the value of the applied voltage [13]

In Figure 5a the dependence of the average amplitude $\bar{A}$ of MP oscillations on the frequency of EF is shown.

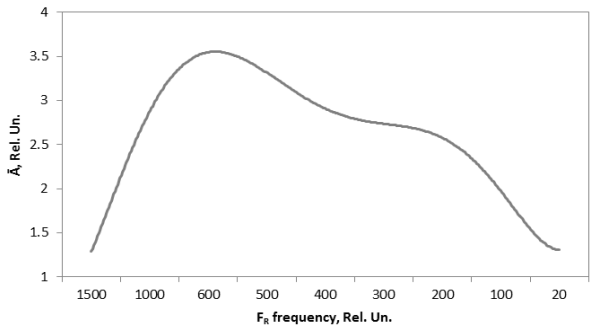


FIGURE 5a. Dependence of the average amplitude, $\bar{A}$ of MP oscillations on the frequency of EF

The corresponding dependence of the average spike height on the frequency of applied voltage is presented in Figure 5b.

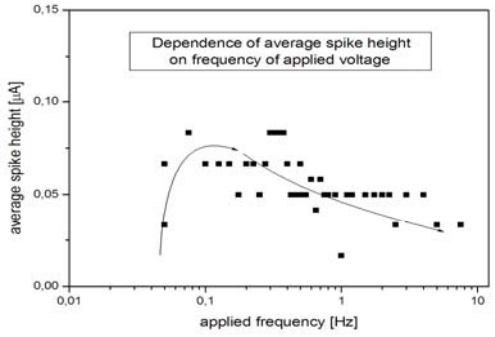


FIGURE 5b.. Dependence on the average spike height on the frequency of applied voltage. Points: measurements, line: drawn to guide the eye [13].

We explain the dependence in the Figure 5b by a "memory effect" which manifests itself in the mechanism of spikes formation: each next spike "remembers" the information about previous spikes. As a result of this effect according to the model there is some optimum frequency of the external exciting factor that provides a maximum average spike height. In the case of our model it means that at higher frequencies of the EF action the model lattice is too disordered after the previous spikes, and the conditions for the synchronization of individual spikes are not favourable. On the other hand, at too low frequencies the model lattice is completely restored after the previous spikes ("forget about the previous spikes"), and thus it hinders the formation of new integrated spike. By the way, the "memory effect" can explain the experimental results showing that the spikes preferentially occur at pronounced, rather equidistant voltages. Preparation of each subsequent spike depends on the system state after the previous spikes, which may be different. The "memory effect" can be also a reason of the fact that the spike spectra are not always reproducible. The influence of previous spikes on the next spike is not the same. The computer experiment with our model shows that the dependence of the maximum value of spike height on the frequency of EF is weak, Figure 6.

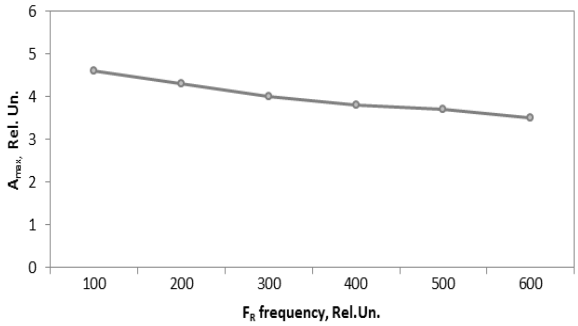


FIGURE 6. Dependence of the maximum value, A_{max} of spike height on the frequency of EF

We revealed that this dependence is determined by the potential of MP interaction. At the same potential, the maximum value of spike height changes weakly. However, for other potentials this maximum value is already different. It means that that for different types of tracks or different materials we can expect different maximum values of spike height.

Figure 7 explains the necessity of a sufficiently large density of MP (and accordingly tracks) to obtain the spike effect. The dependence in this Figure is strong enough. At too small densities of MP the mean time distance between spikes becomes so large that we simply do not see them any longer.

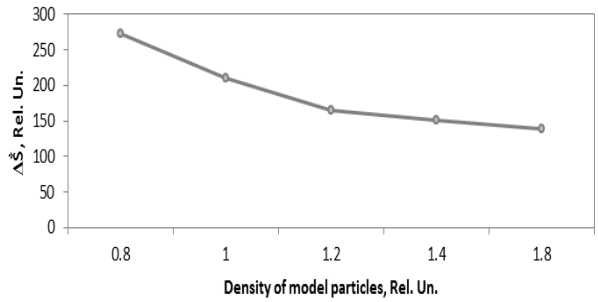


FIGURE 7. Dependence of the mean time interval, ΔŠ between spikes on the density of MP.

4. Conclusions

A computer experiment with the developed model of the pulsating track device showed that the model reflects the main features of the behaviour of such devices. The important result is that the behaviour of the model "device" essentially depends on the potential of MP interaction. The experimental results show that for different types of track-based devices the pulsating effect is determined significantly by the mean distance between tracks (the areal density). We also proceeded from the experimental fact that there are common properties of pulsating track-based devices independent on the materials and physical characteristics of tracks. As a result, the developed model reveals all main features of the real pulsating track devices.

Acknowledgments

Authors are grateful to Prof. David Fuks for useful discussions.

References

[1] Fink D., Petrov A. V., Hoppe K., Fahrner W. R., Papaleo R. M., Berdinsky A. S., Chandra A., Chemseddine A., Zrineh A., Biswas A., Faupel F., Chadderton L. T. (2004) *Nucl. Instr. Meth. Physical Res.* B, **218,** 355–61

[2] Fink D., Petrov A., Fahrner W. R., Hoppe K., Papaleo R. M., Berdinsky A. S., Chandra A., Zrineh A., Chadderton L. T. (2005) *Int. J. Nanoscience,* **4,** 965-73

[3] Fink D,, Chadderton L, T,, Hoppe K,, Fahrner W, R,, Chandra A., Kiv A. (2007) *Nucl. Instr. Meth. Physical Res.* B, **261,** 727–30

[4] Hoppe K,, Fink D,, Fahrner W. R. (2008) *J. Electrochem. Soc.,* **155**, 7-11

[5] Fink D,, Klinkovich I., Bukelman O., Marks R. S., Kiv A., Fuks D., Fahrner W. R., Alfonta L. (2009) *Biosensors* and *Bioelectronics,* **24,** 2702-6

[6] Fink D., Muñoz H. G., Alfonta L. (2012) Highly Sensitive Ion Track-Based Urea Sensing with ion-irradiated polymer foils *Nucl. Instr. Meth.* B, **273,** 164–70

[7] Garcia A. H., Fink D., Muñoz H. G., Alfonta L. (2013) Nuclear track-based biosensors with the enzyme lactase *Biosensor International Conf.* 2013. Kusadaši, Turkey Ed, Samet Acar, 135-44

[8] Mandabi Y., Fink D., Alfonta L. (2013) *Biosensors and Bioelectronics,* **42,** 362–366

[9] Siwy Z. S., Powell M. R., Petrov A., Kalman E., Trautmann C., Eisenberg R. S. (2006) *Nano Letters,* 6, 1729-34

[10] Fink D., Cruz S., Vacik J., Hnatowicz V. (2010) *Rad. Effects Def. Solids,* **165,** 818 –33

[11] Fink D., Vacik J., Hnatowicz V., Muñoz H. G., Alfonta L., Klinkovich I. (2010) Funnel-type etched ion tracks in polymers. *Rad. Effects Def. Solids,* **165,** 343–61

[12] Fink D., Kiv A., Fuks D., Chandra A., Golovanov V., Ivanovskaya M., Khirunenko L. (2007) *Appl. Phys. Lett.,* **91,** 83512-7

[13] Fink D., Kiv A., Fuks D., Munoz H. G., Cruz S. A., Fahrner W. R., Hoppe K. (2011) Computer Modelling and New Technologies, **15**(1), 7-18

[14] Fink D., Cruz S.., Muñoz H. G., Kiv A. (2011) *Rad. Eff. Def. Sol.,* **166**, 373-88

[15] Schlick T. (2002) *Molecular Modeling and Simulation.* Berlin-Heidelberg–New York: Springer-Verlag

[16] Hairer E, Lubich Ch and Wanner G 2003 *Acta Numerica* 399–450 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.155.1628&rep=rep1&type=pdf

[17] Uberuaga B. P., Montalenti F., Germann T. C., Voter A. F. 2005 in *Handbook of Materials Modelling, Part A: Methods,* ed. S. Yip. Berlin: Springer Verlag, 629-48

[18] Frenkel D., Smit B. (2002) *Understanding Molecular Simulation.* San Diego: Academic Press.

[19] Zhen S., Davies G. J. (1983) *Physica status solidi* a, **78**(2), 595–605

[20] Troelsen A. (2010) *Pro C# 2010 and the .NET 4 Platform* 5t^{h} Edition APRESS, p. 1752 http://www.apress.com/9781430225492 10 Nov 2013

[21] Deitel P. J., Deitel H. M., (2010) C# 2010 for Programmers. 4th edition. New Jersey: Prentice Hall 1296 p.

[22] Gutterman A. (2013) *Intermediate Scripting* http://unity3d.com/learn/tutorials/modules/intermediate/scripting Jan 2013

[23] Raynal M. (2013) *Concurrent Programming: Algorithms, Principles, and Foundations.* Berlin: Springer Verlag. 538 p.

Authors

Natalia Mykytenko

Current position, grades: Post-graduate student of the Department of physical and mathematical modelling, South-Ukrainian national pedagogical university after K. D. Ushinskij
University studies: South-Ukrainian national pedagogical university after K. D. Ushinskij.
Scientific interest: She is a specialist in the field of neural networks and classical molecular dynamics. Her works are devoted to modelling of physical processes in track nanodevices.

Dietmar Fink

Current position, grades: Dr., at present he is a guest Professor of the University of Mexico. Earlier he was a Head of laboratory in Hahn-Meitner institute (Berlin), the Editor-in-Chief of International journal “Radiation effects and defects in solids”.
Scientific interest: radiation effects in nanodevices
Experience: one of the founders of track electronics

Arnold Kiv

Current position, grades: Professor-researcher of Department of materials engineering, Ben-Gurion University of the Negev, Head of Department of physical and mathematical modelling, South-Ukrainian national pedagogical university after K. D. Ushinskij.
Scientific interest: nanomaterials and nanoelectronics
Experience: He is an expert in the theory of real structure of solids and physical processes in electronic devices. In last decade, he obtained significant results in the field of nanomaterials, track nanostructures and track electronics.

Nanothinking and nanoeducation: nanoscientific literacy for responsible consumer decision making

T. Lobanova-Shunina[3*] and Yu. Shunin[1, 2]

[1] *Institute of Solid State Physics, University of Latvia, Kengaraga Str. 8, LV-1063 Riga, Latvia*

[2] *Information Systems Manangement Institute, 1 Lomonosov Str, Bld.6, LV-1019, Riga, Latvia*

[3] *Riga Technical Universisty, 1 Lomonosov Str, Bld V, LV-1019, Riga, Latvia*

Received 1 January 2014, www.tsi.lv

Abstract

Consciously or unconsciously, the term 'nanotechnologies' is firmly entering the life of every consumer-citizen of the global community designating both relatively simple nanomaterials and goods that have already entered the market, and very complex technologies that are supposed to change radically the future of mankind. Applications based on today's basic research are expected to form the next industrial revolution. The unique properties of nanotechnology applications suggest potential to solve some of the most pressing social and business challenges, but they come with uncertainties and risks as all new technologies.

Taking advantage of technological progress and preventing adverse side-effects requires analysis, evaluation and guidance to ensure technology is developed in ways that benefits wider consumer society and every individual consumer on the planet. Unfortunately, general public lacks understanding and awareness of the basic properties, and sometimes even the existence of nanotechnologies and their implications linked to the consumption of nanoproducts. Moreover, a generally sceptical attitude among society groups prevails towards new technologies.

The general lack of public knowledge about nanoproducts that are already on the market in a full swing is likely to bring irrational and erroneous, potentially harmful, results. Therefore, modern technology requires educated work-force and responsible consumers and hence imperative for educated population.

Our mission had a focus on introducing changes into the curriculum to eliminate gaps in scientific knowledge of students (as potential consumers, managers and scientists) and to foster an active approach to developing responsible scientific consumption practices and to offer an opportunity for students from a wide range of disciplines to learn about nanoscience and nanotechnology, to explore these questions, and to reflect on the place of new technologies in the spheres of their major and in the global society.

Keywords: Nanotechnologies, responsible scientific consumption, consumer identities, nanoeducation, nanothinking

1. Introduction

"While technology shapes the future, it is people who shape technology and decide what it can and should be used for"
(Kofi Annan)

Modern technology requires educated work force and hence imperative for educated population. The needs of new emerging technologies and a beneficial state of society are compatible in this case. There is no monolithic thing called technology. Rather there are various technologies, which converge or compete to fit into what can be called an ecosystem of technological and societal arrangements. Societal and technological arrangements co-evolve. This co-evolution happens most favourably in an educated, intellectual, and affluent society that is tolerant of change and divergent views. By fostering an educated, intellectual society, it creates conditions that foster responsible moral and social behaviour of the individual and contributes to shaping intellectual humankind [1].

The map of the globe does not show a place called Technoworld, yet in many ways we are already its citizens and consumers. Moreover, as consumers, our lives and behaviours are strongly influenced and shaped by an ever-increasing system of modern technology, transcending national boundaries to create a vast technological synergy grounded in complex applications in industrial production, logistics, electronic communications, agribusiness, medicine, and science. We begin to realise that, like it or not, we have become a part of technoculture.

The industrialized world has developed into a production and consumption global community with a highly advanced level of technologies. This process has led to markedly increased demands for a standard of living and consumption.

However, the current economic crisis has provoked a growing consensus that the 21st century consumer society is on a path that cannot promise its citizens a hope for

* Corresponding author - E-mail: busus@inbox.lv

sustainable future. The prevailing forms of political economy are failing to guarantee the consumers economic stability, preserve ecological resources and services, reduce social inequality, maintain cultural diversity, and protect physical and mental health of citizens. We face related crises of educational, social, cultural and personal sustainability.

On the other hand, current developments in scientific and technological research raise a number of ethical questions comprising responsibility. Areas of research as nanotechnology and biotechnology, regarding food, healthcare and environmental issues, elicit complex and undeniable debates within society today. Scientific-technological research and investments in different areas do not advocate a common good as their overall aim but serve the interests of those who finance the research itself, whilst forcing those who use these products into commercial traps, often without any research and information regarding the effects they may have on consumers' health.

Striking developments of new technologies, particularly nanotechnologies, are finding applications in all spheres of life and producing rapid, systemic, and far-reaching changes in business, government, society and the environment, alongside with the challenges they pose to society. Furthermore, rapid technological changes are forcing organizations to embrace new technologies and change the way they work and interface with suppliers and customers, thus, leading to changes in many behaviour patterns.

However, there is no one-way-fit-all strategy to guarantee a brighter tomorrow for everyone. Nor can the separate efforts of businesses, governments, organizations and individuals cope with the tasks without complementary contributions of others. Yet everyone can benefit from the insights of reasonable research into the nature of change stipulated by the advent of new technologies and innovation.

2. Knowledge management as a means of social change: Who needs nanotechnology education?

According to Petrides & Nodine (2003) knowledge management (KM) brings together three organizational resources – people, processes and technologies – to use and share information more effectively. Knowledge has become the most valuable resource. Prominent technology leaders, nanotechnology boosters, scientists, policy officials, and environmental organizations have raised important questions about nanotechnology's potential economic, social, and environmental implications. However, there is very little knowledge in wider European society about what nanotechnologies are and what impact they might have on how we live. Many experts acknowledge that uncertainties prevail about this.

The central question on nanotechnology education is 'Do we need nanoeducation?' To answer this question, we should first find out who needs nanoeducation? What is the interest in nanoeducation from those who have expressed the need? What kind of education is needed - expertise, skills, level? For what kind of jobs are skills and knowledge of nanotechnology needed?

Nanotechnology has shaken the world and the advanced countries are investing billions of dollars for its R&D and industrial applications. For example, USA cumulative investments in nanotechnology-related research since 2001 now total over 16.5 billion dollars (environmental, health, and safety research since 2005 now total nearly $575 million; education and research on ethical, legal, and other societal dimensions of nanotechnology since 2005 total more than $390 million) (NNI,[2]). Similar amounts are being spent on nanotechnology by Japan, Russia, China and European Union. Nanotechnology has therefore been taken up in these countries as an important national requirement.

The National Science Foundation (NSF) has estimated that by 2015 the world will require about 2.000.000 multidisciplinary trained nano-technologists, including Europe with about 300-400,000 nano-specialists. Therefore, modern technology requires educated work force and responsible consumers and hence imperative for educated population. The needs of new emerging technologies and a beneficial state of consumer society are compatible in this case.

To create a sustainable, democratic, technologically empowered and intellectual global society, higher education has to be at the heart of these processes and play a double role (Reid, et al., [3]). First of all, it has to provide a top-level multidisciplinary education to produce a highly educated workforce. Secondly, to educate the general public by ensuring accessible information that will allow people to better understand what nanotechnology is, how it will be applied, and its implications for the society. Responsible consumption is based on understanding of advantages and threats of new technologies [5, 6].

Thus, higher education must, in one way or another, come to terms with new emerging technologies and identify the paramount place that new technologies have taken in the society. Therefore, the presence of new technologies – as *means, object,* and *context* – in the sphere of contemporary higher education is undeniable (Figure 1).

The European Commission highlights the need to promote the interdisciplinary education and training together with a strong entrepreneurial mindset. It is emphasized that the need for nanotechnologists will not only be confined to the industrial and R & D sectors but will be needed practically in all spheres of life.

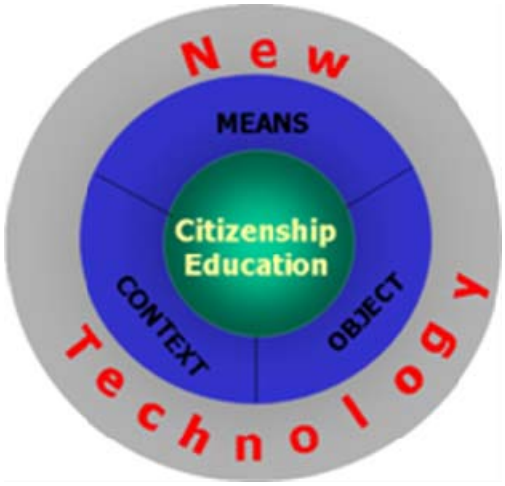


FIGURE 1. New technology as means, object, and context of contemporary higher education

Experts have estimated that marketing of nano-based industrial products will have risen to some 2-3 trillion dollars by 2015, and nanotechnology is going to dominate the socio-economic life of the world for the next 40-50 years [7].

It requires many resources on the consumer side (intellectual, psychological, etc.) to adjust to the rapid changes in the social and business environment by bridging the gap between the change in attitude and the change in behaviour. To a great extent, bridging this attitude-intention behaviour gap is stipulated by the increase of consumer awareness and the raise of the knowledge level.

Therefore, shaping intellectual, responsible consumption practices based on knowledge and awareness is viewed today as the development of civic skills contributing to the sustainable future by enabling people to make their own informed decisions about highly complex technological problems of the day, to take responsibility for their health, their own lives and to contribute to the wellbeing of their communities.

To fulfil the task, it is the job of our higher educational systems of the 21st century to prepare young citizens for the challenges and controversies of the rapidly changing and diverse consumer societies and highly technologically empowered business world. It is the role of higher education to develop skills and values required to enhance democratic life for everyone and to make their informed consumer voices heard in policy decision-making. Democracies need knowledgeable problem solvers and responsible decision makers.

3. Quality education for intellectual youth and sustainable future

The basis of any reflection whether personal or social, rests on an enlightened and critical intellect. Given its ubiquitous nature, nanotechnology is an essential component of responsible citizenship education on the way to intellectual society. It motivates the young adult to shape his thought process, to favour opportunities that refine his critical judgment and allow him to look upon the society of which he is a full member with a clear and constructive eye. He will then be ready to play his role as a knowledgeable citizen and contribute to the ongoing intellectual growth and wellbeing of his community. Sustainability is defined as a long-term maintenance of responsibility, which has environmental, economic and social dimensions and encompasses the concept of responsible management. In its turn, responsible management rests on knowledge and understanding of new technologies and scientific advancements fostering the societal development.

Unfortunately, our previous research on Nanoeducation and Nanothinking has revealed a dramatically low level of basic scientific knowledge and nanotechnology utilitarian value in Latvian students. The research results stimulated the educational component redesign at Information Systems Management University (ISMA), Riga, Latvia. Our mission had a focus on introducing a general nanoeducation course into the curriculum in 2011 to eliminate gaps in scientific knowledge of our students and to foster an active approach to ***Quality Education for Sustainable Future* (QESF).** The general ambition has been to add some non-technical instruction into the curriculum in a way that fits the ISMA Systemic educational model of problem- and project-based learning.

The course is built around active learning methods to promote an active discussion-based approach to developing responsible scientific citizenship and to offer an opportunity for students from a wide range of disciplines, including the natural and social sciences, humanities, business, information technologies, and tourism to learn about nanoscience and nanotechnology, to explore these questions, and to reflect on the place of technology in the spheres of their major and in the global society. We believe that effective and successful development of responsible scientific citizenship depends on a well-balanced integration of the three components:

1) new emerging technologies education,
2) citizenship education, and
3) the humanities education fostering social responsibility.

At the initial stage (the receiving and knowledge levels) the three domains can operate with near independence. Nevertheless, as a person reaches the stage of a high level of techno-scientific proficiency combined with socio-cultural and ethical knowledge (the integration and evaluation levels) the overlap area approaches totality. Scientific knowledge educates citizens about their powers and responsibilities. In this model, the citizen is not a mere consumer of scientific knowledge, but a person whose voice and opinions are heard and valued. Moreover, if the process is strategically targeted, we can view education as a contribution to a scientifically literate, intellectual society (local-European-global)

where young generations are able to take responsibility for its sustainable development (Figure 2).

The QESF course exploits the Systemic approach to the educational process and aims at developing the students' systems thinking, critical thinking, and contextual thinking (learning transfer), as well as organization and communication skills, problem-solving and decision-making abilities, thus, contributing to the integrated skills development.

From the Systemic perspective, an individual, as a social being, educates, self-organizes, and develops his personality through interaction and communication. Inferred from this is the vision that personality is a systemic quality including both biological factors and social formations. Therefore, the Systemic approach to educational process contributes to personality development engaging both biological and social aspects. This process is purposefully organized. From this point of view, the Systemic approach is considered as personality-directed.

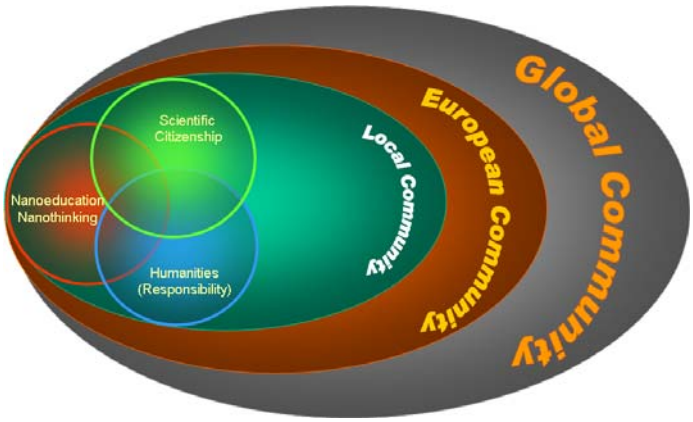


FIGURE 2. Responsible scientific citizenship development based on interdisciplinarity

According to Lev Vygotsky (1991), education can only then be useful if it comes before development. Preceding development, education rests on its achievements, finds the resources for further realization of educational perspectives. Hence is the dynamism of personality development in the educational process. This major principle formulated by Vygotsky means that through interaction and communicative activities, psychic functions (thinking, memory, knowledge accommodation) are formed, as well as social skills, ethic norms, values perception, and self-awareness are developed. From this perspective, the Systemic approach can be viewed as an activity-based approach.

The Systemic education and development promote self-education and self-development, contributing to life-long learning. Inferred from this, we can say that the Systemic approach can be viewed as holistic, leading to the global personality development.

QESF is not a mathematical or technical course. In fact, we spend most of our time on the humanities side of the world, but students get the general knowledge of nanoscience, nanotechnologies and their implications in the society as a result of integration of the humanities with technosciences. The course places emphasis on humanitarian applications of new technologies by focusing on the role of technologies in tackling society's grand challenges such as safety, health and environment. Integration of the humanities with technosciences, envisages the development of scientific competence, providing all the citizens with the abilities to assess new technological and scientific developments, and, thus, be engaged in educated problem-solving forums and responsible decision-making legislatures. The course is to be guided by teams of teachers from a diverse array of disciplines. Science and technology teachers will have to learn humanities and social sciences. Humanities and social science teachers will have to learn sciences and technologies, *promoting interdisciplinarity.*

In the interactive process of interdisciplinary activities, new abilities appear - emergent abilities [8] *that reinforce scientific competence and contribute to its development* (in Figure 3 - arrows inside the circles).

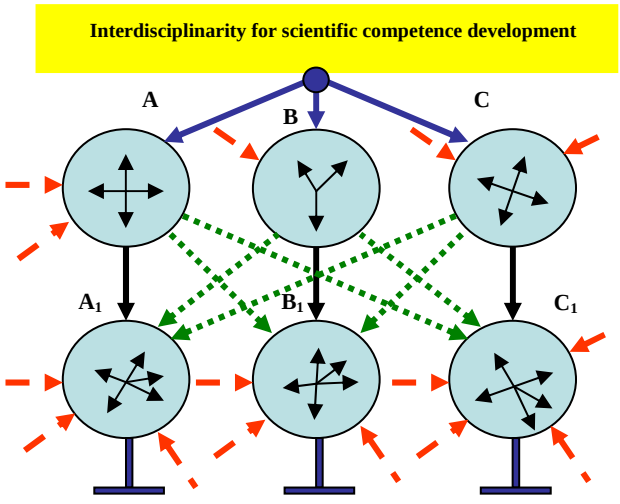


FIGURE 3. Scientific competence development under the systemic approach.

It is necessary to emphasize that the essence of mechanisms showing the emergence of new abilities can be demonstrated only by means of models created within the framework of the systemic approach as these 'new' properties have to be additive.

To a large extent, the course is about connecting disparate questions, concepts, facts, and ideas, and then raising new questions – it is a vital process in this approach to nanothinking because it is a formal way of integrating ideas and communicating.

4. Consumption and wellbeing: Who is consumer?

We commonly think of consumption as something that benefits individuals. We also tend to think of consumption decisions as being made by individuals or families, and not so much by businesses, organizations, governments or other groups. However, in contemporary economies, consumption decisions and consumption benefits are more complicated than this individualistic approach suggests.

From a systemic perspective, consumption is a holistic, cyclic process by which goods and services are used and in the end are disposed of by people (Figure 4).

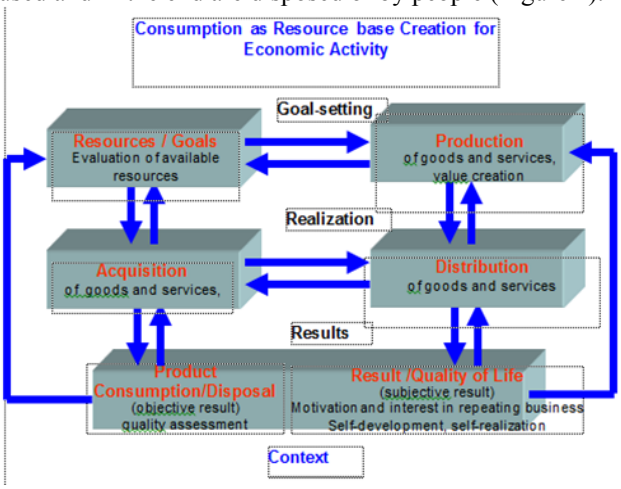


FIGURE 4. Consumption as a resource base creation process for business activity.

Consumption is the final round of the cycle in business activities that starts with an evaluation of available resources and proceeds through production of goods and services, distribution of goods and services, and their acquisition by people or groups. Finally, goods and services themselves come to be used and after their life cycle disposed. The effect of this consumption, including depletion of resources and generation of waste as well as enhancement of people's quality of life and human wellbeing, forms the resource base for the next round of business activity.

Following the concept of a 'cycle' as a unit of the business activities, we can assume that a cycle of production-consumption satisfies a set of important demands:

1) A cycle represents the major relationship 'production – consumption' reflecting the essential certainty of business activity.
2) A cycle is an entire, holistic business act – from the goal to the result, stipulated by feedback ties that ultimately determine the result.
3) A cycle represents a unity of objective and subjective components, which reveals the possibilities for improvement of business activity and product development as well as allows for scientific analysis of the process.
4) A cycle is a system that is open for the influence of the external environment.
5) A cycle represents a model of an objective business reality, since it reflects the basic relationship – the interdependence of production and consumption.

Much of economic discourse, from Adam Smith onward, has assumed that the functioning of an economy is grounded on the final demand for goods and services. As Smith argued, consumption is the sole end and purpose of all production and the welfare of the producer ought to be attended to, only so far as it may be necessary for promoting that of the consumer.

Thus, as the traditional assumption suggests, the consumer is the reason for economic activity and, therefore, for economic theory as well.

Another suggestion is that having satisfied a certain demand, consumers keep the economy going by generating further demand for goods and services. Without this demand, the supply side of the economy would expire. Thus, as a source of demand, the consumer is the trigger that makes the economic system run.

However, people are more than just consumers. Consumption practices most directly address people's living standards or lifestyle goals, which have to do with satisfying needs and getting pleasures through the use of goods and services. The end product derives its value solely from its contribution to the well-being of society and of individual consumers. When everything revolves around the consumer, the question arises: what makes people want to buy and consume?

5. Developing responsible consumer identities through intellectual consumer practices

Marketing professionals strive to influence consumers toward choosing and purchasing a particular brand of their product, at a particular time and place. To succeed, they have to have a clear understanding about what makes people want to buy and consume.

The fast scientific and technological advancements in such areas as neuroscience, genetics, artificial intelligence, biomedical engineering, computer and communications technology, biotechnology, and nanotechnology call for fresh reflections on what it means, in the 21st century, to be a consumer, and for ethical judgments on how we

might shape our intellectual, responsible consumption on the way to sustainable future consumer society.

What is important for a contemporary consumer to be able to cope with all the tasks imposed by new technologies and scientific advancements in order to make independent knowledgeable decisions? But what if those decisions are influenced by producers and aggressive marketing by businesses?

The modern consumer is not an isolated individual making purchases in a vacuum. Rather, we are all part of a contemporary phenomenon that is often referred to as a global consumer society where all people have become increasingly interconnected and interdependent. This aspect could potentially raise greater awareness of the fact that we are all part of a single global community that shares a common consumer identity as a background and destiny.

All of us are human beings and our human identity is the most essential factor that unites all people of the world (Figure 5). At the same time, all of us are consumers: we consume goods and services, and our consumer identity is another major factor making up our commonality.

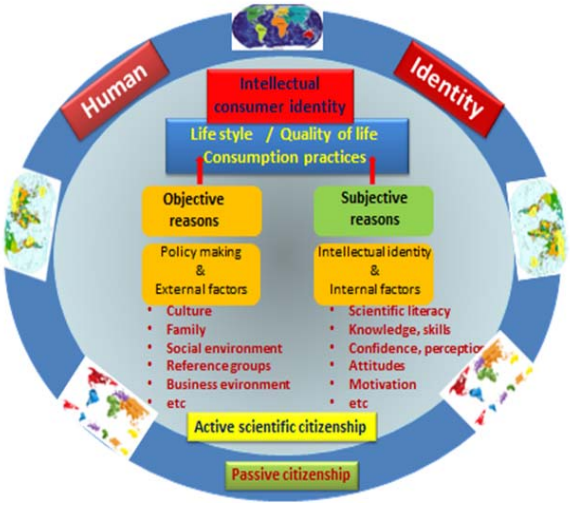


FIGURE 5. Constructing responsible consumer identity through shaping scientific citizenship and intellectual consumption practices

But against the background of this commonality, we develop different consumption practices due to various reasons – botnh objective and subjective. Still, on the subjective side, it is, to a great extent, our intellectual power that shapes our decision-making, our life styles and contributes to the quality of life.

It means that we need to have enough knowledge, skills and confidence in a highly technological business world to effectively construct our responsible consumption practices and develop scientific citizenship with the intention to contribute to innovations, sound business practices and facilitation of responsible and informed policy making to satisfy the requirements of individual consumers as well as contributing to the improvement of the quality of life in general. Education, as a major catalyst, has to help people become effective, scientifically literate, knowledgeable decicion-makers and responsible consumer-citizens. The cost is much greater if it does not.

The rise of consumer culture and the increased scientific literacy can be to a certain extent equated with intellectual citizenship, since a scientifically literate public can better contribute to policy making. Thus, the notions *intellectual citizen* and *responsible consumer* become almost conflated, or, at least, harder to differentiate.

We want our students to leave university with a clear understanding of the political, legal and economic functions of the society they live in, and with the social and moral awareness to thrive in it. Responsible consumption remains a crucial way of organizing people's place in the contemporary consumer society. It will undoubtedly continue to change meaning across space and time in the twenty-first century.

This is why we need to consider the role of new technologies in students' daily lives, in shaping responsible attitude to the consumption practices, and their implications for classroom performances. How closely, for example, should students' worlds outside the classroom match what occurs in the classroom? Why is it important to develop intellectual, responsible attitude to consumption? Intellectual, *responsible consumers* are people showing power of the mind to reason and apply knowledge, who are capable of *choosing* through connecting, of *buying* through thinking, of *consuming* through awareness, of *changing* through understanding.

6. From intellectual youth to intellectual society

In our educational approach we support the idea that instruction should not be separated from practical context and this is the concept that should be maintained as the main classroom philosophy, since skills have not only to be learned, they have to be experienced through the practice of implementation.

Therefore, concurrently with the general nanoeducation course, the quadruple-approach nanoproject ***'From Intellectual Youth to Intellectual Society'*** **(FIYIS)** has been launched that organizes conferences for students as well as the university teachers and general public. Initiated by several professors from Latvia, Israel, and Russia, this project has been designed as an educational supplement featuring the reflections of reputed scientists who propose basic ideas for intellectual growth, while focusing on today's most pressing problems. Among these are Arnold Kiv, Yuri Shunin, Paul Dyachkov, to name a few.

The teaching objectives of the nanoproject FIYIS envision a broad-based integration of technosciences and the humanities at the university level, so that both future technoscientists and humanoscientists develop a common understanding and, possibly, even a common language to deal with complex social, ethical, legal and political

questions arising from the development of nanotechnology and from its convergence with other techno-scientific developments, the answers to which are to be found on the counterbalance of technologies and humanities. Without that counterbalance, society risks scientists without conscience, technicians without taste, and businessmen without responsibility.

The learning context of the NanoProject is subdivided into four basic levels, comprising *Ecological/ Environmental Level, Health and Medical Level, Consumer Goods Level, and Information Communication Technologies Level.*

FIYIS – is an integrated skills project that presupposes the engagement of all students into the research of the Nano-world, irrespective of their major. Thus, the 1st year students tackle the 1st Level tasks: through *simulations, role plays, discussions, and negotiations* they develop their communication and social skills. As a *deliverable* of the project, they might make a nanoproduct presentation: home-related nanoproducts.

The 2nd year students learn to select, process and analyse scientific information dealing with the 2nd Level tasks, thus, developing research-based learning skills, and problem-solving skills.

The 3rd year students acquire all the aforementioned strategies and skills and go a step further – the tasks of the 3rd Level aim at developing not only scientific competence but also at a global personality development of the student *through the experience of learning (attitudinal change – to nanoproducts, to each other, and to the process of learning – i.e. motivation, student awareness, and social scientific responsibility).*

The goal is to produce a pamphlet on key nanotechnologies and nanogoods circa 2015 that may have value to producers, managers and consumers, as well as to future iterations of the class.

There is a special merit in the 3rd Level tasks – that is a high degree of task authenticity, globality, integration with other subjects and involvement of all the aspects of the individual's personality, previous experience and knowledge. Nanothinking and creativity are the factors that link all these elements.

The most common approach, however, as at many other universities, seeks to provide the first-year students with an understanding of the commercial conditions in their subject area, and the project work often involves a market analysis of the particular technical or scientific product that the students are learning how to make.

The second approach provides what might be termed an academic understanding of contextual knowledge. The course provides an introduction to the philosophy of science and technology, and in the project work, the second-year students are encouraged to use these philosophical ideas to consider the ways in which knowledge is produced, or constructed, within their fields of major.

The third type, which we have developed in the new educational program in nanotechnology, can be termed a socio-cultural approach to contextual knowledge. In our lectures, we have introduced the local and the ERASMUS students to the cultural history of science and technology and, in their project work, we have advised the students as to how they might address, and, at best, assess the cultural and ethical implications of the emerging technologies in their fields of major. The approach is valuable since it tends to regard issues of social responsibility and of what we have come to characterize as scientific citizenship.

Normally, the project lasts for one semester, although it can be prolonged for the whole academic year, depending on the tasks and goals. The necessary information is gathered from different sources – course books, extracurricular books, the web, site visits, and interviews with specialists and experts. Students should be encouraged to think of projects as not requiring standard, back-of-the-book answers; rather, different teams undertaking the same project could arrive at different conclusions and deliverables.

A brief pre-assessment is given in the first week of class and two more detailed assessments are given in the last week of class. Several feedback surveys are made during the semester. The assessments and surveys show that the students have found the course valuable and that many of the goals in the syllabus have been met.

Continuous assessment has to be used to gain information about the effectiveness of class discussions and enhancement of students' understanding of the interaction between nanotechnology, individual, and society.

At the end of a project, every student must submit a statement of personal growth - what he/she expected at the initial stages of the project, and what has actually been learnt by the end of the project. The statement of personal growth may incorporate non-traditional objectives, containing reflections on:

- the relevance of the project to the community, city, nation, Europe or the world;
- suggestions for follow-up projects and other activities.

Nanoeducation challenges all students to broaden their horizons and gives them ways of acquiring knowledge of things that shape intellectual society. It fuels their interest as citizens so that they would be curious about the state of current knowledge, regardless of their major. It prepares them to follow the evolution of knowledge and technologies, to be active responsible citizens today and speak knowingly on questions dealing with quality of life within their local communities and the global society.

Not only would the practice of open discussion, problem-solving, decision-making, and statements of personal growth encourage healthy introspection, it would also anchor the scientific and technical disciplines with humanities and social sciences. This is especially important because the exceptional synergy of nanotechnology with other disciplines creates significant

social, legal, ethical and political issues that can be effectively resolved and outspoken only by the intellectual citizenry of an intellectual community. Finally, a gap analysis can be implemented to provide the best way of strategic assessment and planning (see, It allows comparing two series: 1) where we are now and 2) where we want to be in some time in the future, making it easy to identify the gaps in knowledge that need to be closed. For each area giving us a complete picture of the situation we ask two questions:

- Where are we now?
- Where do we need to be in 13 weeks' time?

Actually, we have to answer three questions:

- How are we doing?
- How should we be doing?
- How much do we need to improve? (the gap)

We can then quickly identify where the gaps are and whether things need to improve.

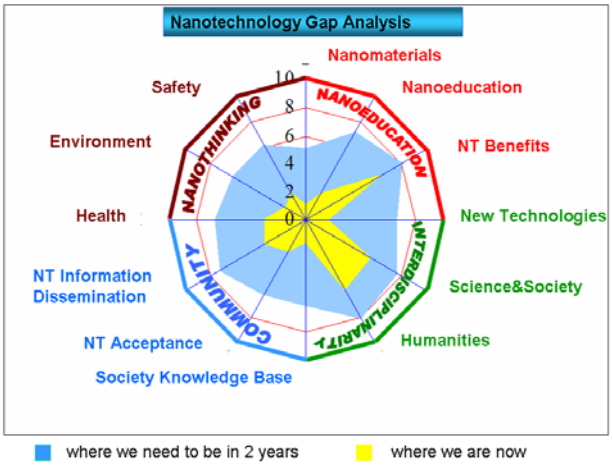


FIGURE 6. Gap analysis of knowledge

7. Towards an open dialogue on the benefits and risks of nanotechnologies with Latvian students

The term 'nanotechnologies' is firmly entering the life of every citizen of the global community designating both relatively simple nanomaterials and goods such as plastic bags or containers, and very complex technologies that are supposed to change radically the future of mankind, such as prosthetic implants that are controlled by the brain and move, feel, and have the sense of touch like real ones; or invisible brain implants that would enhance human memory storing information equal to several big libraries, altering mood and controlling artificial limbs. In fact, nanotechnologies have already entered the market with an extremely wide range of applications comprising food and beverages, food packaging, dental fillers, toothpaste, optics and electronics, clothing, wound dressing, sporting goods, dietary supplements, cosmetics, and many others, but consumers do not seem to notice or care.

Being avid consumers of products, contemporary Latvian youths are very familiar with their wide variety due to the efforts of marketing campaigns, advertising media and their own use of the Internet. However, as they buy and use today's products, they hold no concept of how these products come to exist or how they are made. Overall, general public's knowledge of the production process is relatively limited and unappealing. This lack of knowledge creates a strong demotivating barrier that prevents many potential students from not only entering, but even considering the field. Such a knowledge gap creates a need to educate the students about what constitutes a modern production process enhanced by new technologies, particularly, nanotechnologies, and intellectual consumption.

In this paper, we describe an effort to bridge the technological literacy gap in consumer citizens, currently under way in Latvia. Our previous research on Nanoeducation has revealed a dramatically low level of scientific knowledge in Latvian consumers. The results stimulated the initiation of the project '**A**dopting **I**ntellectual **L**ife **A**pproach' (AILA) at Information Systems Management University (ISMA), Riga, Latvia.

The aim of this research, focusing on the challenges of nanotechnologies in the food and healthcare sectors, is to explore Latvian intellectual responsible consumers: their habits, new technology perceptions, preferences and values. Intellectual, responsible consumption is viewed as an identity project since we will study how Latvian citizens of the recession times construct their identity based on intellectual responsible consumption practices.

The beauty of AILA is that it reaches students from all study programmes – Natural sciences and Humanities, Information technologies and Business studies, Tourism and Management, Law and Environmental design, etc. It is an integrated skills project built around active learning methods to promote an active case-study-, problem-solving- and decision-making-based approach to developing responsible scientific consumption, new emerging identities, and to offer an opportunity for students from a wide range of disciplines to learn about nanotechnologies, to explore their risks and benefits and to reflect on the place of nanotechnologies in their personal life, in their future professional practices, and the modern consumer society.

A questionnaire for the consumer survey comprising a set of questions was compiled based on a Likert Technique or scale and supported by the Consumer Culture Theory by Arnould and Thompson [9]. The main findings of the research are threefold as there were three teaching objectives of investigation: 1) what level of scientific understanding and risk-assessment would be sufficient for consumers; 2) whether consumers are provided with the necessary information; and 3) what more needs to be done in terms of public engagement.

The empirical results have proved a dramatical discrepancy between nanotechnology use in products that

are already on the market and lack of nanotechnology understanding in a modern Latvian consumer society.

The question arises: does higher education today fulfil its role as a major catalyst to provide the necessary knowledge and relevant skills mix for our students to be prepared to join the highly technological global economy to ensure sustainable future for themselves, in the first place?

To find the answer to this question, we initiated a pilot research with the second-year students as an educational supplement, as a part of market analysis of a particular product that the students are learning how to make. The first question addressed to the respondents was whether they considered themselves 'intellectual consumers' (Figure7).

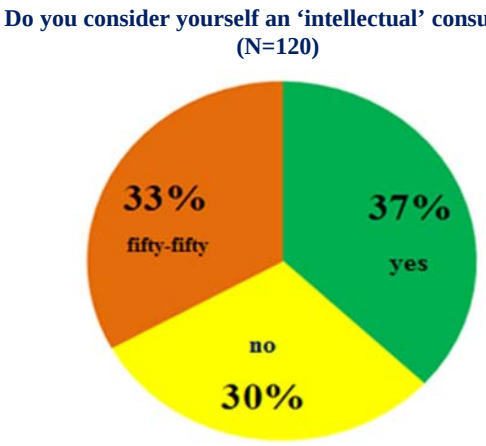


FIGURE 7. The amount of students considering themselves 'intellectual consumers'

As we can see from the pie chart, the answers divided almost evenly among the three groups of students-consumers with a slight shift toward 'intellectual consumer'. Thus, we had to investigate the reasons for 30% of students considering themselves non-intellectual consumers and 33% being doubtful about themselves (Figure 8).

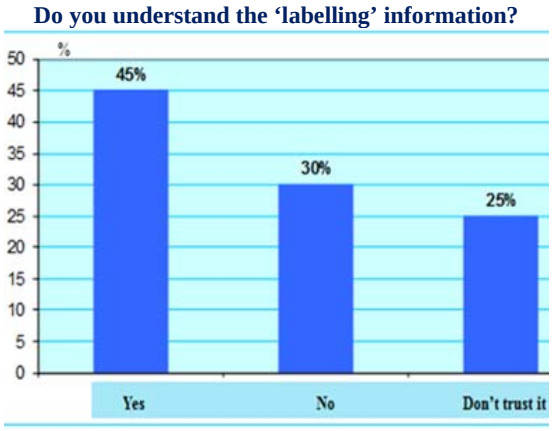


FIGURE 8. Student understanding of the 'labelling' information on products.

As the bar graph demonstrates, the reasons behind such a low self-esteem and doubt are hidden behind the inability to cope with complex scientific inscriptions on the products revealing their ingredients – 30% of students confess not being able to understand all of them. Another 25% of students admit to be very sceptical about information accompanying products because many companies do not label nano ingredients or other unhealthy components for people to choose whether to buy such products or not (Figure 9).

FIGURE 9. The reasons influencing the purchasing decisions of students-consumers

As we can conclude from the pie chart, the major part of consumers is influenced by price in making their purchasing decisions. The second biggest category – 24% of consumers put their health on the first place, which is very reasonable.

Still, judging from the bar graph (Figure 10), the plurality – 62% – have no idea about products containing nano-ingredients, hence we can assume that they cannot make responsible decisions concerning their health.

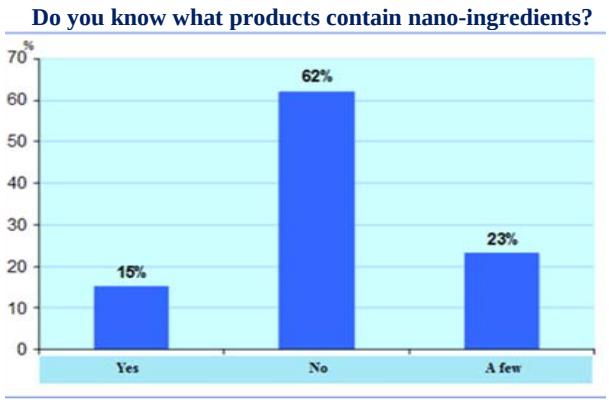


FIGURE 10. Students-consumers' awareness of the products containing nano-ingredients

Only 15% of students-consumers say that they know what products contain nano-ingredients. However, if we sum up 15% and 23% (of shy students, perhaps) we will get that 38% who consider themselves intellectual consumers.

Among students-consumers who make an initial assessment of nanotechnologies, the plurality think the risks and benefits will be about equal, and the votes are divided about evenly between benefits and risks – 60% and 40% respectively (Figure 11).

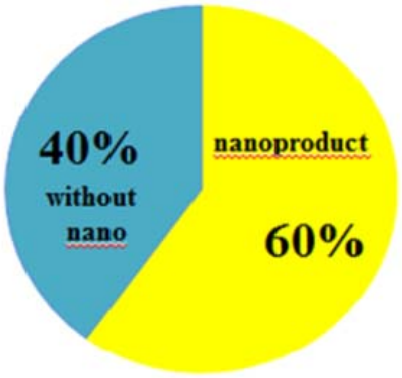


FIGURE 11. Student preferences of buying products with or without nano-ingredients *before* a general course on NTs.

However, when potential risks and benefits are outlined in the general introductory course into nanotechnologies, the greatest shift is toward the need of additional information – 74% or risk – 19% (Figure 12).

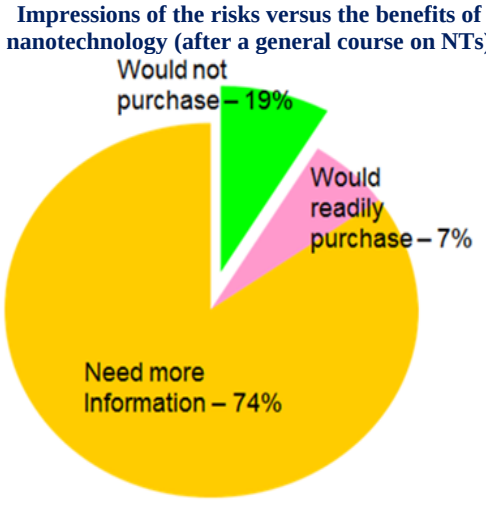


FIGURE 12. Student preferences of buying products with or without nano-ingredients *after* a general course on NTs

As we can see, life itself imposes the necessity of knowledge and education to make informed choices and responsible consumer decision making, thus, ultimately becoming intellectual consumers.

Since Nanotechnology is starting to play an extremely important role in the socio-economic development of all countries for the foreseeable future, it is imperative for higher education that emphasis be placed on producing a properly educated, qualified and trained specialists and consumers that can cater for the future of the society they live in.

It is at this crossroads and the questions raised by the scientific and technological advancements of our contemporary world, that we need to transform the society by educating responsible decision-makers and contributing to the construction of new emerging identities, shaping scientific responsible consumption practices and lifestyles.

Societal and technological arrangements co-evolve. This co-evolution happens most favourably in an educated, intellectual, and responsible society that is tolerant of change and divergent views. 'By fostering an educated, intellectual society it creates conditions that foster responsible moral and social behaviour of the consumers and contributes to shaping intellectual responsible mankind' (Gomez-Mejia, [10]). Our choices today will determine what kind of tomorrow is likely to come.

8. Conclusions

There is now a growing consensus that the global society of the 21st century is on a way that is not sustainable. The contemporary forms of political economy are failing to conserve ecological resources and services, to guarantee economic stability, to maintain cultural diversity, to ensure environmental security, and to protect people's physical and mental health. We face corresponding crises of ecological, economic, social, cultural and personal sustainability.

Nanotechnology is expected to radically alter the human condition within a short span of time, probably not exceeding two decades. Human cultures, however, do not change at the same rapid pace. Technoscientists as well as humanoscientists begin to ponder and predict the parameters of possible social, environmental, ethical and legal changes to emerge in the first two decades of the new millennium. At the same time, the preparation of future leaders and engaged citizenry to cope with unpredicted changes has obviously to begin at schools and universities.

Most of today's higher educational institutions are awash in technology but the outcomes for students remain little changed from 20 years ago. The problems are not in our technology but in our universities. The reasons are due to our attitudes toward education - how we underfund it, mismanage it, politicize it and disempower it in our culture. We have to dare shake education out of its two-centuries-old inaction/inertia. New technologies and education are inseparable. Involving the general public into decision making is a key element of social learning for sustainability.

Education for sustainable development should develop knowledge and understanding of the social, economic and environmental dimensions. Addressing the social dimension clearly involves citizenship education based on knowledge and understanding of the new technologies fostering the societal development.

The European Union is stimulating the development of nanoscience education in universities to address complex issues and to solve multidisciplinary problems, in general.

From a practical stance, nanotechnology is widely considered to be 'the next big thing' and is well worth learning more about in order to get a knowledgeable understanding of what is nanotechnology? Is it all hype? Is it dangerous? This essential knowledge will not be developed by chance, but by strategic, targeted teaching.
On another level, at the intersection of technology and society, there is a new angle to think of some timeless issues. Is nanotechnology good? What is progress? How much risk are we ready to take? Why should we care about the societal implications of nanotechnology? These are profoundly important questions the answers to which are to be found on the counterbalance of technologies and humanities. Without that counterbalance, society risks scientists without conscience, technicians without taste, and businessmen without responsibility.

Values reflect and shape the ongoing social development. Understanding the impact of a new technology on society is vital to ensuring that development takes place in a responsible manner. Scientific knowledge is expected to play an important role in educating citizens about their powers and responsibilities. In this case, the citizen is not a mere consumer, but a person whose opinions are valued.

A democracy needs an educated citizenry. What does an educated citizenry in a technological age look like? To participate in a democracy influenced by technology not only do citizens need to know how to understand the multiple perspectives that they encounter, they need to feel an obligation to explore multiple perspectives to fully understand the society they live in and make informed decisions.

When we do not pay close attention to the decisions we make, when we fail to educate ourselves about the major issues of the day, when we choose not to make our voices and opinions heard, that is when citizenship and democracy breaks down.

A new interdisciplinary course, developed at ISMA, places an emphasis on humanitarian applications of new technologies by focusing on the role of nanotechnologies in tackling society's grand challenges such as safety, health and environment. Our hypothesis is that this new approach to teaching about technologies will engage and inspire students who have typically been turned off by the traditional educational experience [11]. Additionally, we believe that this course will better prepare a new generation of specialists to address major societal problems in the future maintaining at the same time an awareness of political, economic, ethical and social constraints on technologies.

Acknowledgements

Authors are thankful to Prof. Arnold Kiv and Prof. Stefano Bellucci for fruitful discussions.

References

[1] Langdon W. (2010) The Whale and the Reactor: A Search for Limits in an Age of High Technology. Chicago: University of Chicago Press
[2] National Nanotechnology Initiative-NNI. (2012) http://www.nano.gov/
[3] Reid A., Jensen B. B, Nikel J., Simovska V. (2008) Participation and Learning: Developing Perspectives on Education and the Environment, Health and Sustainability. *Participation and Learning. Perspectives on Education and the Environment, Health and Sustainability,* ed A. Reid, B. B. Jensen, J. Nikel and V. Simovska Dordrecht: Springer, 1–18
[4] Petrides L. A., Nodine T. R. (2003) Knowledge management in education: Defining the landscape http://iskme.path.net/kmeducation.pdf
[5] Lobanova-Shunina T., Shunin Yu. (2011) *Computer Modelling and New Technologies,* **15**(1), 58-68
[6] Lobanova-Shunina T, Shunin Yu. (2011) Nanothinking as a Concept of Citizenship Education for the 21st Century Europe *Europe's Future: citizenship in a changing world,* ed. P Cunningham and N. Fretwell. London: CiCe, 345-55
[7] Meyer M. (2007) *Scientometrics,* **70**(3), 779-810.
[8] Lobanova T., Shunin Yu. (2009) Language as a social behaviour: integration of language, communicative competence and identity development, *Human Rights and Citizenship Educatio*n, ed. A. Ross. London: CiCe 306-12
[9] Arnould E. J., Thompson C. J. (2005) *Journal of Consumer Research,* **31**(4), 868–82
[10] Gomez-Mejia L. R, Balkin D. B, Cardy R. L. (2008) *Management: People, Performance, Change,* New York: Hill/Irwin
[11] Lobanova-Shunina T., Shunin Yu. (2012) Nanoeducation for developing responsible scientific citizenship and creating a sustainable intellectual global *community Creating Communities: Local, National and Global,* ed. P. Cunningham and N. Fretwell. London: CiCe, 316-26

Authors

Tamara Lobanova-Shunina, Rezekne, Latvia

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 nanothinking in the EU FP7 Project CACOMEL (2010 to 2014), special interests concern the systemic approach to nanosystems applications.
Publications: 55
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.

Yuri N. Shunin, born 06.03.1951, Riga, Latvia

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.
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).
Scientific interest: His current research activities concern nanophysics, nanoelectronics, nanodevices, nanomaterials, nanotechnologies, nanorisks, nanoeducation, and nanothinking
Publications: about 470, 1 book with Springer
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

Supplies transportation planning in power grid urgent repair based on hierarchical genetic algorithm

J. Zhou[1,2] *, W. Q. Ma[3] and J. Y. Miao[4]

[1]*School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China*

[2]*School of Economics and Management, North China Electric Power University, Beijing 102206, China*

[3]*Department of Information Technology, Shanxi Yangquan Electric Power Corporation, Yangquan 045000, China*

[4]*Huada Tianyuan (Beijing) Electric Power Technology CO., LTD, Beijing 102206, China,*

Received 1 January 2014, www.tsi.lv

Abstract

To optimize supplies transportation solutions of power grid, a hierarchical genetic algorithm is put forward. Double hierarchical objective function is set by weighting, that is, when the precondition of inequality constraint is met, objective solution is to find out the shortest supply time and if the solution is also within the time constraint, the final solution would be the least supply cost, otherwise, it would still be the shortest supply time. The solution for optimal supplies transportation scheme at the least cost would be worked out by iteration of genetic algorithm. Compared with single objective genetic algorithm in simulation, hierarchical genetic algorithm is proved more effective and superior to decrease economic loss of accidents.

Keywords: Hierarchical Genetic Algorithm, Time constraint, Transportation Cost, Transportation Time, Economic Loss

1. Introduction

After a power grid accident, the first important thing is to quickly make emergency supplies allocated to the destination, thus to decrease economic loss. Whenever a great disaster occurs, one supply warehouse could often hardly meet the demand for emergency supplies so that to cause the problem of multi-site to rescue. In Literatures [1, 2], the problem of supplies scheduling within the shortest time from the least supplies sites are discussed in terms of strictly mathematical proof. In Literatures [3, 4, 5], the optimal path multi-material in electric power scheduling is worked out by Dijkstra algorithm and an overall optimization on planning is realized by establishing an independent scheduling system. In Literature [6], emergency supply scheduling at single accident site is studied by taking the least supply sites and the shortest emergency responding time as the objective functions. In Literature [7], under the condition of time constraint at single supply site and multiple accident sites, the scheduling problem of single type supplies is discussed, and a model for emergency supply scheduling is proposed to aim at the shortest responding time.

So far, researches on urgent repair mostly focus on solutions for optimal path. However, this paper tries to find out a solution to decrease economic loss by analysing the problems of multi- sites of supply & demand, and scheduling. Taking the minimal economic loss as the objective, optimal planning in power grid emergency repair is elicited by iteration of genetic algorithm based on a double-hierarchical model of the shortest transportation time and minimal transportation cost.

2. Modelling

To simplify model, assumptions are made below.

a) Impacts by traffic jams, traffic light and so on would be ignored, that is, the consumed time is a constant from a supply site to a demand site.

b) Quantities for different demanded supplies are integers, that is, there are no requirements to combine specific quantity and special type supplies at fault sites.

c) There is no restriction on vehicles at each supply site, that is, arbitrary transportation at a supply site is allowed.

Assumed that there are m power grid supply sites and $P=\{P_1,P_2,...P_m\}$, where there are l kinds of supplies and $S=\{S_1,S_2,...S_l\}$, supplies would be distributed to n fault sites and there is $Q=\{Q_1,Q_2,...Q_n\}$. Given that the demand amount for supplies k at the i-th supply site is A_{ki}, where there are $k\in(1,l)$ and $i\in(1,m)$, while the demand amount for supplies k at the j-th supply site is B_{kj}, where there are $k\in(1,l)$ and $j\in(1,n)$. Assumed that

* Corresponding author. Tel: +86-010-6177-2698; fax: +86-010-6177-2696;
E-mail: zhoujing108@ncepu.edu.cn

the consumed time from the supply site i to the fault site j is t_{ij}, where there are $i \in (1,m)$ and $j \in (1,n)$. c_{ij} is the unit cost by transport supplies from supply site P_i to Q_j. The purpose of supplies transportation planning in urgent repair is to determine the supply amount from each supply site to the demand site, i.e., to find the solver of x_{ijk}, which is the amount of supplies S_k transported from supply site P_i to demand site Q_j. D_{ijk} represents the activity of supplies transportation from supply site P_i to demand site Q_j, when x_{ijk} is more than 0, there is $D_{ijk}=1$, otherwise there is $D_{ijk}=0$. For an arbitrary k, if there is $D_{ijk}=1$, there is $D_{ij}=1$, that is, the supplies transportation from supply site i to demand site j occur, where there are $i \in (1,m)$, $j \in (1,n)$, and $k \in (1,l)$.

According to assumptions and descriptions above, the model would be established as below.

$$\min\left\{\lambda_1 \times \max\left(D_{ij}t_{ij}\right) + \lambda_2 \sum c_{ij}x_{ij}\right\}. \tag{1}$$

There are three inequality constraints below.

$$\forall k, \sum_i A_{ki} \geq \sum_j B_{kj}, \tag{2}$$

$$\forall j, \forall k, \sum_i x_{ijk} \geq B_{kj}, \tag{3}$$

$$\forall i, \forall k, \sum_i x_{ijk} \geq A_{kj}. \tag{4}$$

In formula (1), when the real transportation time is less than the critical time t_m, there are $\lambda_1 = 0$ and $\lambda_2 = 1$; on the contrary, then λ_1 and λ_2 would be calculated by normalization of weights from practical data. Besides, $D_{ij}t_{ij}$ represents the consumed time from site i to site j in this algorithm. Formula (2) shows that the total amount of supplies S_k from all supply sites could meet the demand at fault site. Formula (3) shows that the amount of supplies S_k sent to fault site Q_j has actually met the demand. Formula (4) shows that the amount of supplies S_k from supply site P_i is less than its reserve [8].

3. A double hierarchical planning on supplies transportation

After a power grid accident, the most important thing for material transportation is to quickly make emergency supplies allocated to the destination, thus to lower economic loss to minimum. In practice, t_{kij}, the consumed time from resource site to demand site, is always assumed to be a constant, then the time taken to handle the power grid fault is $T_s = \max\{t_{kij}\} + t_k$, where t_k is the time to repair. Therefore, when there is $T_s \in (0, T_{\min})$, there is $L(t)=0$, where $L(t)$ is the function of economic loss to time, and when there is $T_s \in (T_{\min}, T_{\max})$, there is $L(t) = k(T_{\max} - T_{\min})$ with a linear relationship. When T_s is longer than $T_{\max}$, the most severe economic loss $L_{\max}$ would be there. Usually, the needed time to repair would be taken as the critical time T_r to judge if economic loss has been caused or not since once supplies are sent to demand site, the needed time to repair T_k is only related to the skills of rescue team without obvious variation. However, the supplies transportation time T_m would change with the road actual situation and the supplies scheduling & planning, so to tremendously influence T_r.

The objective of single hierarchical algorithm is to directly calculate the shortest path or the consumed time based on the shortest path. However, not only an analysis of the shortest path but also a quantitative analysis, where if the consumed time meets the requirement of shortest time at minimal economic loss, should be done in practice. Under this condition, when there is a solution for the shortest path, the objectives would change into the minimal transportation cost and minimal number of rescue; when there is no a solution, the objective would be the shortest transportation time due to the trivial distribution cost comparing with the economic loss by fault itself [9].

Therefore, double hierarchical indicators should be established, on the upper hierarchy are economy-effectiveness-type indicators to ensure supplies transportation within the shortest time; on the lower hierarchy are cost-type indicators which will realize the minimum of objective function only after meeting the condition of the shortest transportation time on the upper hierarchy as shown in formula (1). In practice, great amount of supplies always are badly needed at the spot of fault site within a short time so that those distant and time-consuming transportations are necessary but undesirable while considering the limitation of time. When the time constraint is not met, c_{ij} of transportation cost for this route would be substituted with a figure much bigger than it is possible in practice, other transportation costs for other route would not change [10]. That is, supplies transportation planning would change under different time constraint.

4. Hierarchical Planning Based on Genetic Algorithm

Supply sites would be different with the different limitation on transportation time at fault sites. When the premise of time constraint is met, the optimal and

minimal solution of objective function could be worked out by data comparison.

4.1. CODING

Since the purpose of supplies planning is to find out the optimal transportation plan, codes could be expressed as $(x_{111}, x_{112},..., x_{121}x_{122},..., x_{211}x_{212},..., x_{mnl})$, where x_{mnl} represents the amount of supplies l from supply site m to demand site n, P is the set of supply sites, and Q is the set of demand sites.

4.2. INITIAL POPULATION'S ESTABLISHMENT

$tempx_{ijk}$ represents the amount of supplies k from the practical supply site i to fault site j and $tempA_{ik}$ is the left one of supplies k at site i. According to statistical data, there is a limitation on the optimal shortest time, within which the rescue could be implemented at a minimal loss or beyond which the loss would raise with the increase of the consumed time. Given that the optimal shortest time is t_m, there would be two situations: The first one is that supplies in reserve could meet the demand at fault site for each supply sites where there is $t_{ij} \le t_m$; The second one is that supplies only at those supply sites where there is $t_{ij} > t_m$ could meet the demand at fault site. The process is described step by step and shown in Figure 1.

Step 1: To judge if $\sum_i x_{ijk} \ge B_{kj}$ is false at fault site j, or the process goes to Step 5.

Step 2: For specific fault site j, the set E of supply sites, where $t_{ij} \le t_m$ is true, would be established, then to judge if $temp\sum_i\sum_k x_{ijk} < \sum_j B_{kj}$ is true, where $temp\sum_i\sum_k x_{ijk}$ represents the total amount of supplies k at fault site j, or the process goes to Step 4.

Step 3: For the above site j, supply site r, where there is $t_{rj} > t_m$, would be push into the serial R. Then those t_{rj} while $r \in R$ are sorted in ascending order that the maximal amount of supplies as $x_{rjk} = A_{rk}$ would be added to the total amount of supplies as $B'_{kj} = B'_{kj} + A_{rk}$ until there $B'_{kj} = \sum_i x_{ijk} \ge B_{kj}$ is true. Finally, t_{ij} would be assigned to t_m.

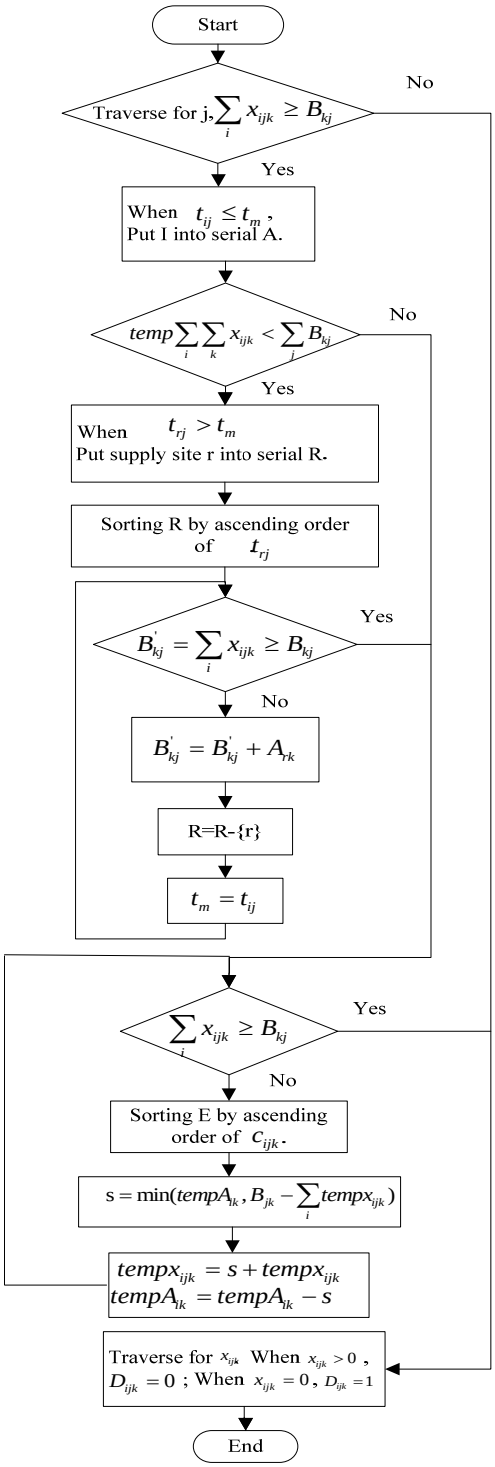


FIGURE 1. Flow chart of double hierarchical genetic algorithm model

Step 4: To judge if $\sum_i x_{ijk} \ge B_{kj}$ is true, as for supplies k, the set E would be sorted according to the ascending order of c_{ijk}, the amount s of supplies from i to j is $s = \min\left(tempA_{ik}, B_{jk} - \sum_i tempx_{ijk}\right)$ according to the

ascending order, where $\sum_i x_{ijk}$ are those values, which meet conditions in above steps. If there are $tempx_{ijk} = s + tempx_{ijk}$ and $tempA_{ik} = tempA_{ik} - s$, the process will redo step 4.

Step 5: As for site j , when x_{ijk} is true, there it is $D_{ijk} = 1$ at site i , or there it is $D_{ijk} = 0$ and the process goes to the end.

4.3. FITNESS FUNCTION

Since the objective function is designed as the minimum, the fitness function would be designed as $F(x) = 1\frac{1}{f(x)}$, where $f(x)$ is the minimal objective function.

4.4. SELECTION, CROSSOVER AND MUTATION

Selection operator would be conducted by means of wheel selection. The locations of supply sites and fault site are fixed for initial codes, thus crossover operator would be conducted in terms of crossover from a supply site to another one. To show the general features of the algorithm, the probabilities of crossover and mutation are expressed with random number as $P_c = random(0.5,1)$ and $P_r = random(0.01,0.1)$ respectively. To meet the limiting conditions above and to avoid infeasible solutions, the process is described below.

Step 1. According to the result of random number, chromosome could be selected to pair. If the pairing result is $random[0,1] < P_c$, where P_c is the probability of crossover, pairing by crossover operation would be conducted, or mutation operation would be done.

Step 2. In terms of stochastic method, the random number of w within the interval of $[1,l]$ would be got and then the crossover transform on the practical amount of supplies w would be conducted.

Step3. An entity is randomly selected from the population, and judgment that if $random[0,1] < P_r$ is true or not would be made, where P_r is the mutation probability. When the condition is met, mutation operation would be conducted: Firstly, the practical amount of transported supplies to demand site j is set as 0. Then, the supplies planning would be reset for demand site j according to the process of population initialization.

5. Case Studies

In this paper, the genetic algorithm tool kit of Matlab is applied. For a same case of emergency scheduling, simulations would be conducted by algorithms of the shortest time-consuming method and double hierarchical model. By comparison of those results, the feasibility of double hierarchical model could be test. Given that there are 3 types of supplies at 4 supply sites offered to 4 fault sites which need rescue in an emergency repair scheduling of power grid, the reserve amount of supplies, the demand amount of supplies and the transportation time would be given by the method of random number to simulate all possible sudden situations in practice. The relationship between supply sites and demand sites is shown in Figure 2 below.

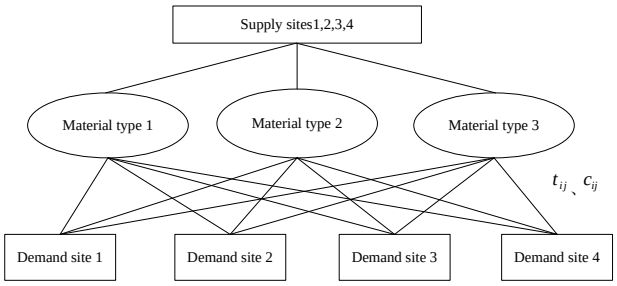


FIGURE 2. Relationship between supply sites and demand sites in emergency scheduling system

Transportation Time t_{ij} is set as a random number in the interval of [6, 20], Transportation Cost c_{ij} is assigned as a random number in the interval of [2, 10], Supplies Reserve A_{ki} at supply sites is a random number in the interval of [15, 45] and demand Supplies B_{kj} at demand sites is a random number in the interval of [10, 40]. If there is $\sum_i A_{ki} < \sum_i B_{kj}$, then random numbers should be reset. The scale of population is 50 with the crossover possibility of $P_c = 0.85$ and the mutation possibility of $P_r = 0.07$, and the limitation of transportation time T_m should be set by considering the severity at fault sites and the practical transportation time from supply site i to demand site j . Therefore, the limitation time would be given by experts, which is $T_m = 10$ in this paper. If there is a solution under this limitation, there are $\lambda_1 = 0$ and $\lambda_2 = 1$, or the objective would shift into the shortest transportation time. Since there are differences between transportation cost and transportation time on both magnitude and significance, coefficients should be set after fully considering experts' advice [11, 12], which are $\lambda_1 = 10/(10+1)$ and $\lambda_2 = 1/(10+1)$ here. In terms of Matlab, the above process would work out two sets of data on the amount of supplies at both supply sites and demand sites, consumed time, and cost as shown in Table 1 below.

TABLE 1. The amount of supplies, consumed time and cost

Parameter	Situation1	Situation 2
A_{ik}	35,27,30 26,37,18 24,31,39 26,30,38	25,39,34 27,40,23 26,23,25 30,28,38
B_{jk}	20,30,34 33,28,19 18,35,29 35,27,23	24,30,14 14,16,22 21,12,15 13,26,17
t_{ij}	12,14,11,8 7,11,12,18 13,12,6,15 12,10,11,12	10,7,10,6 13,5,8,10 9,12,5,10 8,14,8,15
c_{ij}	2,3,5,8 4,5,7,2 1,4,6,7 2,5,7,3	3,5,7,9 6,8,6,4 7,6,7,3 7,6,5,10

In situation 1, when transportation time is longer than T_m at the least loss, the results of simulation are listed in Table 2. Under the same preconditions, the optimal solutions for two algorithms are the same, i.e., the shortest transportation time is 12 and the transportation cost is $\sum_{i,j} c_{ij}\left(\sum_k x_{kij}\right)=1853$, which means that two algorithms are equally accurate in this situation.

TABLE 2. Optimal solutions for two algorithms in situation 1

Algorithms	Shortest transportation time	Optimal solution x_{kij}
Genetic algorithm	12	$x_{141}=35\ x_{142}=27\ x_{143}=23$ $x_{211}=20\ x_{212}=30\ x_{213}=18$ $x_{331}=18\ x_{332}=31\ x_{333}=29$ $x_{421}=26\ x_{422}=28\ x_{423}=19$ $x_{221}=6\ x_{432}=2\ x_{413}=3$ $x_{232}=2\quad x_{321}=1$
Double hierarchical genetic algorithm	12	$x_{141}=35\ x_{142}=27\ x_{143}=23$ $x_{211}=20\ x_{212}=30\ x_{213}=18$ $x_{331}=18\ x_{332}=31\ x_{333}=29$ $x_{421}=26\ x_{422}=28\ x_{423}=19$ $x_{221}=6\ x_{432}=2\ x_{413}=3$ $x_{232}=2\quad x_{321}=1$

In situation 2, when the limiting condition of $T<T_m$ is met, the objective function would be solved to find out the least transportation cost and the results of simulation are listed in Table 3. When the shortest transportation time for two algorithms are all within the limitation of 10, there are $C_1=\sum_{i,j} c_{ij}\left(\sum_k x_{kij}\right)=1732$ for Genetic algorithm and $C_2=875$ for double hierarchical genetic algorithm, which shows that double hierarchical genetic algorithm is obviously superior to genetic algorithm on expenditures.

TABLE 3. Optimal solutions for two algorithms in situation 2

Algorithms	Shortest transportation time	Optimal solution x_{kij}
Genetic algorithm	9	$x_{221}=14\ x_{222}=16\ x_{223}=22$ $x_{331}=21\ x_{332}=12\ x_{333}=15$ $x_{141}=13\ x_{142}=26\ x_{143}=17$ $x_{411}=24\ x_{412}=28\ x_{413}=14$ $x_{312}=2$
Double hierarchical genetic algorithm	10	$x_{111}=24\ x_{112}=30\ x_{113}=14$ $x_{341}=13\ x_{342}=23\ x_{343}=17$ $x_{242}=3\ x_{221}=14\ x_{222}=16$ $x_{223}=22\ x_{431}=21\ x_{432}=12$ $x_{433}=15$

It shows in the two tables above that when transportation time is longer than the time constraint at the least loss, the results of the shortest transportation time are the same for two algorithms, and when the condition of minimal time constraint is met, the value of objective function's solution for double hierarchical genetic algorithm decreases since the objective function takes the least transportation cost as the main factor, while for genetic algorithm, the shortest transportation time is still the objective of the solution with amount of transportation cost. In comparison, double hierarchical genetic algorithm is better than single objective genetic algorithm.

6. Conclusions

According to the nature of power grid supplies planning, a hierarchical genetic algorithm is proposed to make a planning for emergency scheduling and path selection in this paper. Based on genetic algorithm and with double hierarchical objective function set by weights, this algorithm could iterate to find out the optimal solution for supplies scheduling & planning at the least loss after an accident when the inequality constraints are met. In a study case, it shows that the hierarchical genetic algorithm could effectively decrease economic loss of an accident under the limitation of time. We believe this method could be widely used in power grid emergency supplies scheduling.

Acknowledgments

This work is supported by the Beijing Higher Education Young Elite Teacher Project (Grant No. YETP0704) and the National Natural Science Foundation of China (Grant No. 71271084).

References

[1] Liu C. L., He J. M., Shi J. J. (2001) The Study on Optimal Model for a Kind of Emergency Material Dispatch Problem. *Chinese Journal of Management Science*, **9**(3), 29-36, *(in Chinese)*
[2] He J. M., Liu C. L., You H. Y. (2001) Selection of Multi-depot in Emergency Systems, *Systems Engineering-theory & Practice* **21**(11), 89-93 *(in Chinese)*
[3] Ye P. Y., Du H. J., Shen X 2006 Application of the Dijkstra algorithm in the best repairing path, *Relay* **34**(12), 39-41 *(in Chinese)*
[4] Zhang M. Q., Xu M., Qin W. T. (2008) Research of the best repair path based on an improved particle swarm optimization in power communication network, *Science Technology and Engineering* **22**(8), 5990-5 *(in Chinese)*
[5] Wu Q., Teng H., Wang K. F. (2005) Structure of electrical repairs real-time scheduling system based on GPRS/GPS/GIS. *Relay* **33**(17) ,70-3 *(in Chinese)*
[6] Chen D. Q., Zheng W. C., Ding X., Lou W. J., YE C., Zhang Q. (2009) Emergency materials distribution model with time-varying demand and supply constraints. *Logistics Technology* **19**(7), 90-2 (*in Chinese*)
[7] Chai X. R., Wang R. J. (2010) Research of emergent material dispatching algorithm based on multi-depot and multi-materia. *Computer Engineering and Applications* **46**(6), 224-6 *(in Chinese)*
[8] Zhao M., Song X. Y., Dong J., WANG S. J. (2012) Using Genetic Algorithm to Solve Scheduling Optimization Problem of Emergency Supplies. *Journal of Shenyang Jianzhu University* **28**(5), 944-8 *(in Chinese)*
[9] Han Q. (2007) Bilevel Programming Model for One Type of Emergency Material Dispatch Problem and Its Algorithm. *Chinese Journal of Management Science*, **15**(z1), 716 *(in Chinese)*
[10] Reis R., Ritt M., Buriol L. S., Resende M. G. (2011) *International Transactions in Operational Research*, **18**(3), 401-23
[11] Sipahi S., Timor M. (2010) *Management Decision*, **48**(5), 775-808
[12] Wang H. K., Shi Q., Wang G. Y., Wang H. D. (2012) The Model of Sequencing Decision-making of Battlefield Damage Repair for Equipment based on Information Entropy and TOPSIS Method. *Journal of Ordnance Engineering College* **24**(5), 5-9 (*in Chinese*)

Authors

Jing Zhou, born in Gansu, December 2, 1978

Current position, grades: Lecturer, School of Control and Computer Engineering, North China Electric Power University. Ph.D. student of North China Electric Power University
University studies: North China Electric Power University, in the School of Control and Computer Engineering (1998-2005)
Scientific interest: knowledge engineering, modelling and control of complex system, security analysis and risk control technology of power system
Publications: 10 papers, 2 patents

Weiqing Ma, born in Shanxi, April 1, 1965

Current position, grades: Master degree, Senior Engineer and A-class Expert on Engineering Technology of Shanxi Electrical Technology Institute
University studies: Taiyuan University of Technology, in the College of Computer Science and Technology (1984-1988)
Scientific interest: information management engineering and decision support system
Publications: 8 papers, 5 patents

Jinya Miao, born in Jiangsu, June 5, 1990

Current position, grades: Master of electrical engineering, mainly engaged in power system operation analysis, control and management
University studies: North China Electric Power University, in the School of Electrical & Electronic Engineering (2008-2013)
Scientific interest: power system operation analysis and control, operation and management of power grid
Publications: 4 papers

Study of maneuvering target tracking algorithm based on Kalman filter and ANFIS

Zengqiang Ma[1,2] *, Yacong Zheng[1], Sha Zhong[1], Xingxing Zou[1]

[1] *School of Electrical and Electronics Engineering, Shijiazhuang Tiedao University, 17 Northeast, Second Inner Ring, Shijiazhuang, China*

[2] *Key Laboratory of Traffic Safety and Control in Hebei, 17 Northeast, Second Inner Ring, Shijiazhuang, China*

Received 25 December 2013, www.tsi.lv

Abstract

Although Kalman filtering algorithm has been widely used in the maneuvering target tracking, conventional Kalman filtering algorithm always fails to track the maneuvering target as the target changes its movement state suddenly. In order to overcome its disadvantages, an improved Kalman filtering algorithm that based on the adaptive neural fuzzy inference system (ANFIS) is proposed in this paper. In the improved algorithm, the covariance matrix of Kalman residual is gainer and the measurement noise covariance can be updated in real-time by ANFIS module. Finally, the comparison and analysis of the experiment results between the original Kalman filtering algorithm and the improved one has been carried out. The experiment results show that the tracking error is obviously reduced and the accuracy is significantly boosted after the original Kalman filtering algorithm was substituted by the improved one.

Keywords: maneuvering Target tracking, Kalman filter, Adaptive Neuro-Fuzzy Inference System (ANFIS)

1. Introduction

Kalman filtering algorithm was put forward by Kalman in literature [1] in 1960, and it was promoted to continuous system together with R S Bucy in 1961 [2]. It is not only very common in engineering applications, but also has a good performance in moving target tracking. Kalman filtering algorithm plays a good role in the non-motor vehicle target tracking. However, when the target is in the maneuvering motion, the tracking model cannot keep pace with the actual motion model changes and the target state estimation will deviate from the true status, thus Kalman filter will diverge and the tracking accuracy will drop rapidly. In view of the defect of the original Kalman filtering algorithm, scholars have put forward various improved methods, such as Federal Kalman Filter, Extend Kalman Filter (EKF), and Adaptive Kalman Filter, etc., and among them EKF is used widely very much. However, a precondition of EKF is that it must follow the assumption of Gaussian process, which is not always satisfied in the actual application systems [3]. In the EKF, parameters cannot be adjusted easily. Further, more, poor robustness and high computing complexity are also its disadvantages. Because the EKF is usually used in the linear system, which results in a suboptimal application of the recursive estimation of the standard Kalman Filter [4, 5], tracking error will be too large to be used in the maneuvering target tracking.

The Adaptive Neural Fuzzy Inference System (ANFIS) [6] is proposed by Jyh-Shing Roger Jang in 1993. It inherits the advantages of fuzzy logic and neural network. At the same time, ANFIS has discard not only the disadvantages of the fuzzy inference system, such as strong dependence on expert experience and poor self-learning ability, but also that of the neural network, such as lack of training samples and poor universality of its system architecture. Therefore, it can not only adjust its structure adaptively according to the operation environment of the system, but also obtain the membership function by blended learning rule.

Because of the advantages of ANFIS, an improved algorithm, which is based on the original Kalman Filter and ANFIS, is proposed to update the measurement noise covariance in target tracking system in this paper [7]. The simulation results show that the improved algorithm has higher accuracy in maneuvering target tracking than the original one.

2. Analysis of the Original Kalman Filtering Algorithm

2.1. DEFECT OF THE ORIGINAL ALGORITHM

The target status equation and measurement equation are defined as follows:

$$X(k)\Phi X(k-1)+W(k-1), \tag{1}$$

$$Z(k)=HX(k)+V(k). \tag{2}$$

* Corresponding author. Tel: +86-0311-87939226; fax: +86-0311-87936465;
E-mail: mzqlunwen@126.com

In the status equation (1), X is a state vector and Φ is a transition matrix of the system. In the measurement equation (2), Z is a measurement vector and H is a measurement matrix of the system. $W(k)$ and $V(k)$ are zero-mean Gaussian noises, representing system and measurement noise respectively with covariance Q and R.

Kalman filtering algorithm consists of two steps named prediction and measurement update. Based on the linear system above, the prediction step is derived as follows:

$$\hat{X}(k|k-1) = \Phi\hat{X}(k-1|k-1), \tag{3}$$

$$P(k|k-1) = \Phi P(k-1|k-1)\Phi^T + Q(k-1). \tag{4}$$

In formula (3), $\hat{X}(k|k-1)$ is a state vector prediction, $\hat{X}(k-1|k-1)$ is a state vector estimation of $X(k-1)$ at time instant $k-1$. Formula (4) is an error covariance matrix of the prediction state, here, $P(k-1|k-1)$ is an error covariance matrix of $X(k-1)$ at time instant $k-1$.

The measurement update step is carried out by the following equations:

$$\tilde{Z}(k) = Z(k) - H\hat{X}(k|k-1), \tag{5}$$

$$S(k) = HP(k|k-1)H^T + R(k), \tag{6}$$

$$K(k) = P(k|k-1)H^T S^{-1}(k), \tag{7}$$

$$\hat{X}(k|k) = \hat{X}(k|k-1) + K(k)\tilde{Z}(k), \tag{8}$$

$$P(k) = (I - K(k)H)P(k|k-1). \tag{9}$$

Kalman residual is denoted as $\tilde{Z}(k)$ and described as formula (5). The theoretical covariance matrix of it is showed as formula (6). In addition, Kalman gain is depicted in formula (7). $\hat{X}(k|k)$ is a state estimation, $P(k)$ is an estimation error covariance matrix. Compared with the formula (6), the observed covariance matrix of Kalman residual $\hat{C}_2(k)$ can be obtained by the following formulas:

$$E\left[\tilde{Z}(k+1)\right] = \bar{\tilde{Z}}(k+1), \tag{10}$$

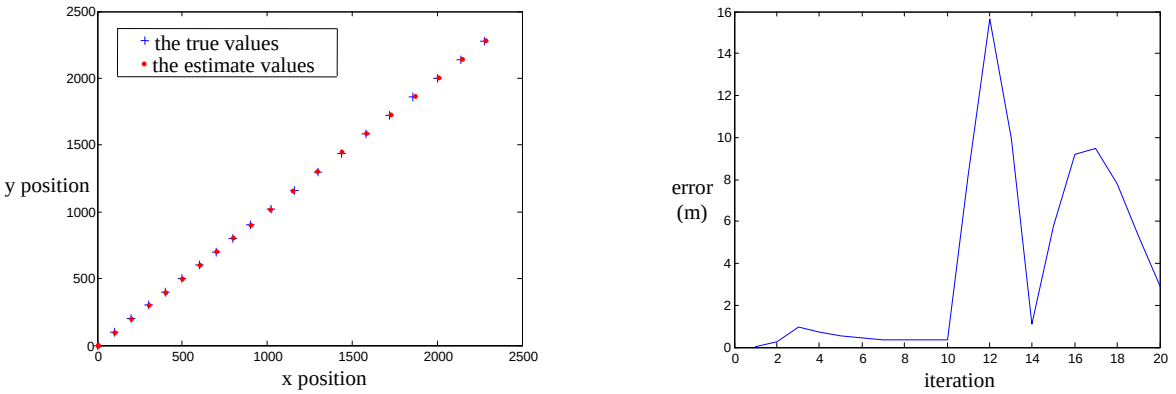
$$E = \left[\left(\tilde{Z}(j+1) - \bar{\tilde{Z}}(j+1)\right)\left(\tilde{Z}(k+1) - \bar{\tilde{Z}}(k+1)\right)^T\right] = \begin{cases} \hat{C}_2(k+1), & k = j \\ 0, & k \neq j \end{cases}. \tag{11}$$

It is can be known from formula (6) and (11) that there are differences between the observed covariance matrix and the theoretical value of Kalman residual. When the target moves at a constant speed, the mean value of Kalman residual is zero. However, this value will change when the target have accelerations. Then, the observed covariance matrix of Kalman residual will not be consistent with the theoretical one [8]. In addition, it is inferred that the covariance matrix of $\tilde{Z}(k)$ is affected by the value of measurement noise covariance as indicated in formula (6). In order to remedy the flaws of the original method mentioned above, the improved algorithm is put forward to enhance tracking accuracy.

2.2. EXPERIMENT RESULT OF THE ORIGINAL ALGORITHM

In this paper, the original algorithm is applied to the simulation experiment of moving target tracking. Supposing that the initial position of the target is $(x, y) = (0m, 0m)$, the initial velocity is $(v_x, v_y) = (50m/s, 50m/s)$. The sample time is $T = 2s$. In the 10th sampling period $(t = 20s)$ an instantaneous acceleration $(a_x, a_y) = (10m/s, 10m/s)$ will be attached to the target, lasting $2s$. Tracking results are showed in Figure 1.

Figure1 (b) shows that when the target moves at a uniform speed, the tracking accuracy is high, but the tracking error increased instantly and the tracking effect reduced suddenly after the speed of the moving target has been changed. From the defects of the original algorithm analyzed in 2.1, it is known that as the speed of the moving target changes, the mean value of the Kalman residual will change correspondingly. Moreover, it will cause the difference between the observed covariance matrix of the Kalman residual and the theoretical one, which will increase the filtering error. This is the main reason of Kalman filter performance drop rapidly. This paper presents a new method that based on ANFIS to improve the original Kalman filtering algorithm. The improved algorithm can update the measurement noise covariance matrix in real-time, then keep the observed value of the Kalman residual covariance consistent with the theoretical one, and finally, make the target tracking experiment getting a better result.



a) b)

FIGURE 1. Experiment results of the original Kalman filtering algorithm in target tracking: a) Tracking displacement, b) Tracking error.

3. The Improvement of Kalman Filtering Algorithm

To overcome the limitation that the original algorithm, an improved IMM algorithm, which combining the original algorithm with an ANFIS module, is presented in this paper.

3.1. PRINCIPLE OF THE IMPROVED ALGORITHM

From the analysis of original algorithm and the simulation result showed in Figure 1 (b) we can know that the target maneuvering will cause the observed covariance inconformity with the theoretical value of the Kalman residual, the tracking accuracy will decline rapidly and Kalman filter will diverge. Formula (6) shows the measurement noise covariance can influence theoretical one, then the filtering effect will be affected also. So updating the measurement noise covariance can keep the observed covariance of the Kalman residual consistent with the theoretical one, and the tracking accuracy will be improved finally. Therefore, an improved Kalman filtering algorithm, in which the measurement noise covariance is updated in real-time by ANFIS, is proposed. The flow chart is showed as Figure 2

In the improved algorithm, the theoretical covariance of Kalman residual is defined as:

$$S(k) = HP(k|k-1)H^T + \hat{R}(k), \tag{12}$$

where $\hat{R}(k)$ takes the place of measurement noise covariance signed $R(k)$ in the original algorithm. The symbol $\hat{R}(k)$ is an estimate value of $R(k)$, which is updated by ANFIS in real-time.

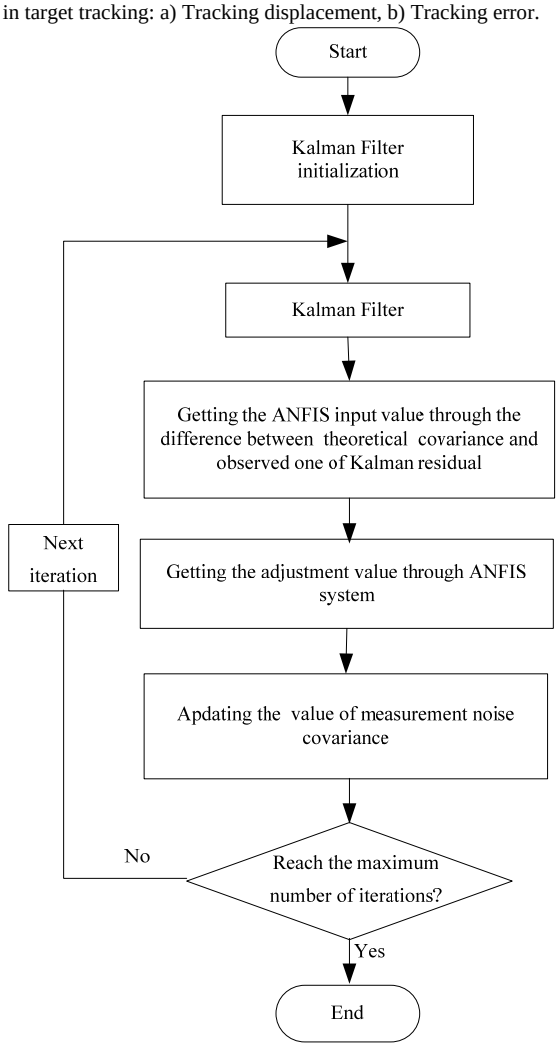


FIGURE 2. Flow chart of the improved algorithm.

Applying Adaptive [9] Neural-Fuzzy Inference System (ANFIS) can produce fuzzy rules and adjust membership functions automatically based on data without experience of experts. Based on those advantages of the ANFIS module mentioned above, the tracking accuracy is evidently

increased in the improved algorithm by updating the value of system noise covariance in real-time.

The principle of the improved IMM algorithm is shown in Figure 3. $D(k)$ is the input parameter of the added ANFIS module, $\Delta R(k)$ will be adjusted to adapt to the target movement state variation.

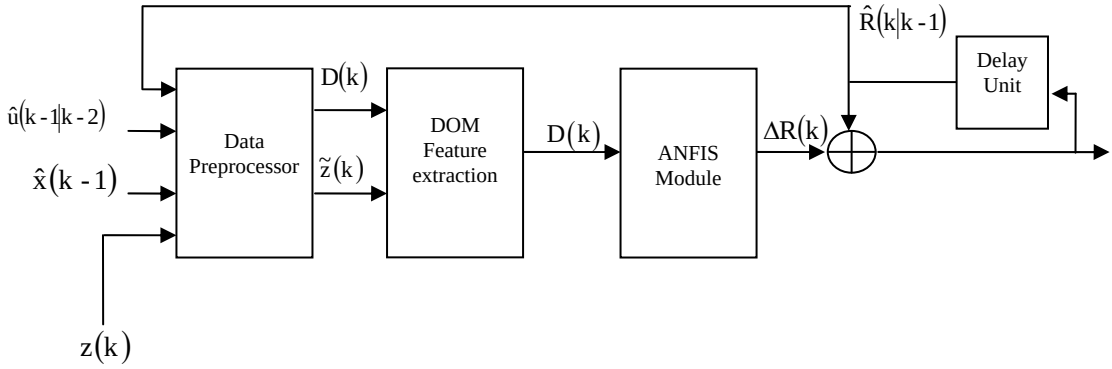


FIGURE 3. Principle of the improved algorithm.

3.2. DESIGN OF ANFIS SYSTEM

It can be known that the observed covariance disagrees with the theoretical value of the Kalman residual when the speed of the moving target changes in the original algorithm. In order to overcome the defect, the improved algorithm updates the measurement noise covariance in real-time based on $D(k)$ by ANFIS to ensure that the two values of the Kalman residual covariance are consistent with each other.

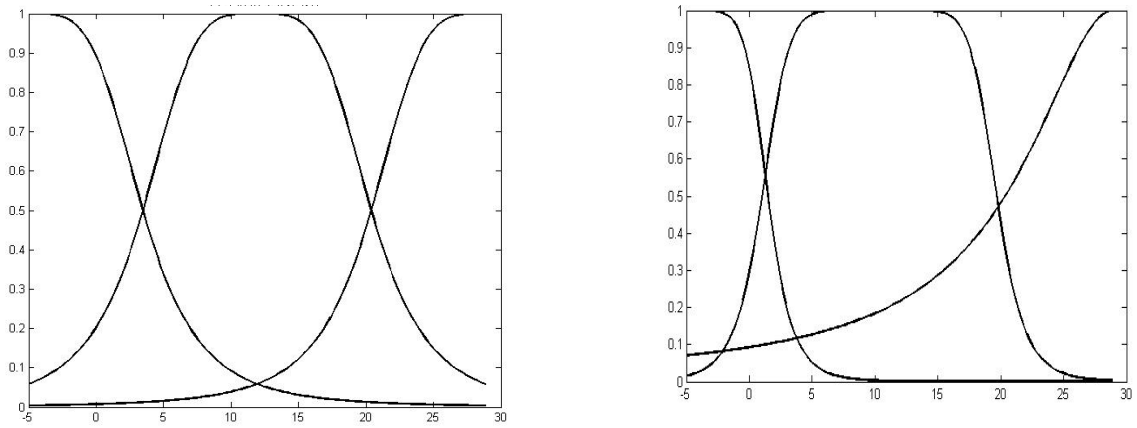
The variable $D(k)$ is defined as:

$$D(k) = S(k) - \hat{C}_2(k) \quad (13).$$

Measurement noise covariance can be updated as follows:

$$\hat{R}(k) = R(k) + \Delta R(k). \quad (14)$$

The input value is mapped into fuzzy sets in the universe of discourse. They are labelled as: $N = Negative$, $Z = Zero$, $P = Positive$. In this paper, the generalized bell shaped function has been chosen as the membership function to characterize fuzzy sets. The initial and final membership functions of the input parameter are showed as Figure 4:



a) b)

FIGURE 4. Initial and final generalized bell shaped membership: a) Initial membership functions, b) Final membership functions.

According to formulas (6) (11) and (13), the value of $D(k)$ is calculated. The value of $\Delta R(k)$ is get by expert experience. In this way, a set of $D(k)$ and $\Delta R(k)$ are obtained. Finally, with these parameters defined above, the fuzzy rules [10] are expressed as follows:

Rule 1 *if* $D(k)$ is *N*, then $\Delta R(k)$ is *N*.

Rule 2 *if* $D(k)$ is *Z*, then $\Delta R(k)$ is *Z*.

Rule 3 *if* $D(k)$ is *P*, then $\Delta R(k)$ is *P*.

Two data sets named $D(k)$ and $\Delta R(k)$ are generated by fuzzy rules mentioned above, which corresponding to the input and output parameters respectively for training of ANFIS. The parameters of ANFIS are showed as TABLE 1.

TABLE 1. ANFIS parameters

ANFIS info:
Number of nodes:16
Number of linear parameters:6
Number of nonlinear parameters:9
Total number of parameters:15
Number of training data pairs:110
Number of checking data pairs:0
Number of fuzzy rules:3

Because the fuzzy inference system based on Takagi-Sugeno is simple in algorithm and convenient to be realized, it is used in the two ANFIS modules. The generated ANFIS structure is showed as Figure 5.

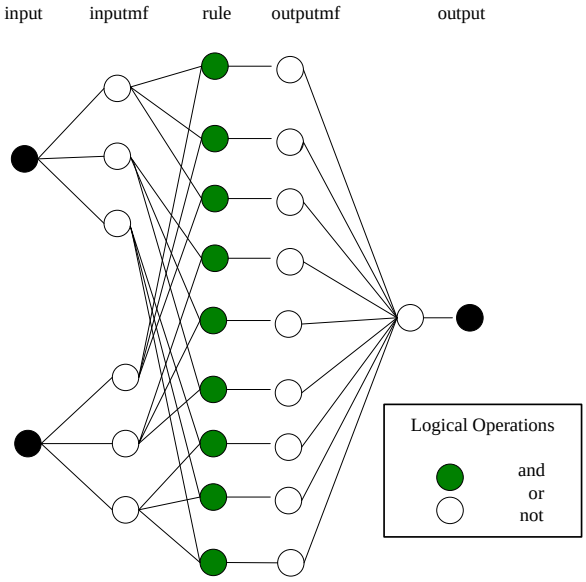


FIGURE 5. Structure of ANFIS

4. Experiments Comparison of the Two Algorithms

4.1. EXPERIMENT RESULT COMPARISON

The improved Kalman filtering algorithm is applied to the simulation experiment of tracking maneuvering target. Both the initial state and the motion state of the target are the same as the previous experiment, that the initial position of the target is $(x, y) = (0m, 0m)$, the initial velocity is $(v_x, v_y) = (50m/s, 50m/s)$. An instantaneous acceleration $(a_x, a_y) = (10m/s, 10m/s)$ will be attached to the target at time instant $t = 20s$, lasting $2s$. And the sample time is also $T = 2s$. Figure 6 is the result of target tracking error. In this figure, the experiment errors of the improved Kalman filtering algorithm and the traditional one are compared.

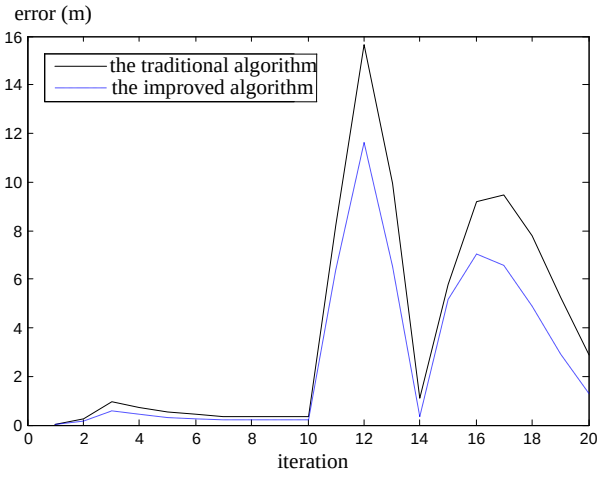


FIGURE 6. Comparison of the estimation error curves

The simulation result shows that the error of original Kalman filtering algorithm is very big and is reduced effectively by the improved algorithm when the speed of the moving target changes. So the loss of the moving target have been prevented owe to the improved algorithm.

4.2. EXPERIMENT ERROR ANALYSIS

The experiment error of the algorithms before and after improved are given in Table 2. It shows that the error of the improved algorithm has obviously decreased compared to the original algorithm from 11th iterations. In addition, the error of the improved algorithm is 3/5 of the original one in general. The reason of improved tracking accuracy is that the measurement noise covariance is compensated in real-time, and the observed covariance of Kalman residual is corrected when the acceleration of target is change.

TABLE 2 Experiment errors of the two algorithms

Iteration rounds	Error before improved (a)	Error after improved (b)	b/a
1	0.04535	0.02636	0.5812569
2	0.25667	0.14942	0.5821483
3	0.97354	0.58923	0.6052448
4	0.7369	0.43314	0.5877867
5	0.55952	0.32587	0.5824099
6	0.4435	0.25693	0.5793236
7	0.37398	0.21707	0.5804321
8	0.34496	0.20273	0.5876913
9	0.34515	0.20525	0.594669
10	0.35542	0.21134	0.5946204
11	8.21708	6.37459	0.7757731
12	15.64758	11.60113	0.7414009
13	9.93892	6.5259	0.6566005
14	1.1217	0.34532	0.3078541
15	5.7885	5.16354	0.8920342
16	9.2082	7.04156	0.7647054
17	9.48966	6.59441	0.6949048
18	7.77816	4.90665	0.630824
19	5.27737	2.94023	0.5571393
20	2.87997	1.29959	0.4512512
Average	3.9891065	2.770513	0.6174035

5. Conclusions

Firstly, the principle of the original Kalman filtering algorithm for maneuvering target tracking and its defects are introduced in this paper. Then, an improved Kalman filtering algorithm based on ANFIS is proposed. In the improved algorithm, because the value of measurement noise covariance is updated in real time by ANFIS module, the theoretical value of the Kalman residual covariance can keep space with the observed one. Finally, the simulation experiments are carried out and the tracking error are quantitatively analysed. The results show that the accuracy of the maneuvering target tracking has been increased obviously, when the original Kalman filtering algorithm is replaced with the improved one.

However, the improved algorithm still has some flaws. For example, the tracking accuracy of the improved algorithm will be influenced greatly by the parameters of ANFIS. So, further research is needed to gain the rules that how the parameters influence the tracking accuracy in the future.

Acknowledgments

This research is supported by Scientific Research Fund of Hebei Provincial Education Department (ZH2012063), National Training Programs of Innovation and Entrepreneurship for Undergraduates (201210107005) and National Natural Science Foundation of China (11372199, 51208318).

References

[1] Kalman R. E. A. (1960) New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering Series D,* **82,** 35-45

[2] Kalman R. E., Bucy R. S. (1961) New Results in Linear Filtering and Prediction Theory. *Transactions of the ASME Journal of Basic Engineering. Series D,* **83,** 95-107

[3] Rigatos G. G. (2012) Nonlinear Kalman Filters and Particle Filters for integrated navigation of unmanned aerial vehicles. *Robotics and Autonomous Systems*, **60,** 978-95

[4] Rigatos G. G. (2008) Particle Filtering for state estimation in industrial robotic systems. *Journal of Systems and Control Engineering.* **222**(6), 437–55

[5] Rigatos G. G., Tzafestas S. G. (2007) Extended Kalman Filtering for Fuzzy Modelling and Multi-Sensor Fusion. *Mathematical and Computer Modelling of Dynamical Systems*,**13,** 251–66

[6] Jang J. S. R. (1993) ANFIS:Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems,Man and Cybernetics*, **23**(3), 665-85

[7] Xu Y. K., Liang X. G., Jia X. H. (2013) Adaptive maneuvering target state estimation algorithm based on ANFIS. *Systems Engineering and Electronic*, **35**(2), 250-55

[8] Ramazan H., Mohammad T., Ali N. M. (2011) A novel adaptive neuro-fuzzy unscented kalman filter for slam. *International Journal of Humanoid Robotics*, **8**(1), 223-43

[9] Zhangsong Shi, Pixu Zhang, Shen Li, Rui Wang (2012) An adaptive tracking algorithm of maneuvering target. *Advanced Materials Research: Manufacturing Science and Technology,* **383-390,** 2179-83

[10] Escamilla A. P. J., Mort N. (2002) Multi-sensor data fusion architecture based on adaptive Kalman filters and fuzzy logic performance assessment. *Proceedings of the Fifth International Conference on Information Fusion,* **2,** 1542-49

Authors

Zengqiang Ma, born 04.04.1975, Hebei province of China

Current position, grades: associate professor
University studies: Ph.D. degree was earned in major of vehicle engineering, Beijing Jiaotong University in 2011
Scientific interest: digital signal processing
Publications: He has published many papers in recent years and more than ten of them were included in the international Engineering Index (EI)

Yacong Zheng, born 20.06.1987, Hebei province of China

Current position, grades: graduate student in the Shijiazhuang Tiedao University
University studies: Bachelor degree was earned in major of electronic and information engineering, Yanshan University in 2011
Scientific interest: digital signal processing
Publications: She has published two papers

Sha Zhong, born 10.09.1991, Hunan province of China

Current position, grades: graduate student in the Shijiazhuang Tiedao University
University studies: Bachelor degree was earned in major of measuring and control instrument, Shijiazhuang Tiedao University in 2013
Scientific interest: digital signal processing

Xingxing Zou, born 10.06.1993, Hunan province of China

Current position, grades: undergraduate student in the Shijiazhuang Tiedao University
University studies: Bachelor degree will be earned in major of electronic and information engineering, Shijiazhuang Tiedao University in 2014
Scientific interest: digital signal processing

The property verification methods of complex stochastic system based on directed graph

Ji Mingyu [1, 2 *], Li Yanmei [2], Chen Zhiyuan [2]

[1]*College of Information and Computer Engineering, Northeast Forestry University*

Harbin, 150040, China

[2]*College of Computer Science and Technology, Harbin Engineering University*

Harbin 150001, China

Received 1 January 2014, www.tsi.lv

Abstract

With the increasing complexity of computer hardware and software systems, how to ensure system accuracy and reliability becomes an increasingly pressing issue. The quantitative verification of multiple until formula property has important practical significance in the field of biology. In this paper, for particular probability reward model, we give the detailed analysis of properties verification methods of the multiple until formula with transition step and transition reward constraints based on the weighted directed graph. At last, the example analysis is given. The theoretical analysis and example result show that the feasibility and validity of the method.

Keywords: Probabilistic system; Model checking; Multiple until formula; Directed graph

1. Introduction

Model checking is a formal verification technique with higher degree of automation and is an efficient automatic detecting mean as it can provide the counterexample information when the properties are violated. Model checking has obvious advantages compared with other formal verification means.

In recent years, along with the application range of traditional model checking technology used for qualitative verification continues to be expanded, the properties quantitative validation technology based on complex parametric probabilistic model become one of the research topics that many experts and scholars interested, and gradually become new hotspot in trusted system verification field.

In the process of model checking, probabilistic computation tree logic formula and continuous stochastic logic formula are generally used to describe the property of system. On these logic formula, multiple until formula can describe the periodic oscillations changes of biological species and other important characteristics in the field of systems biology, which makes high-performance verification and counterexample analysis about this kind of property become an open research topic [1, 2].

In the literature [3], stratified continuous time Markov chain model is used to explain the verification process of multiple until formula property with time boundaries, the corresponding verification algorithm and example analysis process are proposed by constructing synchronous automata model of formula property automaton and continuous time Markov chain model.

Literature [4] proposes an algorithm on counterexample generation for model checking probabilistic timed automata based on the weighted directed graph.

In the literature [5], the semantic representation of the probabilistic timed automata was gave by Markov decision processes, and the until formula counterexample generation algorithm and representation method were proposed.

Although domestic and foreign scholars have made a lot of works in the field of until formula properties verification, however most of the existing research are oriented to the until formula with time constraint and rarely study until formula properties verification with the reward constraints. In view of this, the paper will give the detailed analysis of properties verification methods of multiple until formula with transition step and transition reward constraints in Markov reward model.

The rest of paper is organized as follows: Section 2 presents the necessary background on probability model with reward parameter. In Section 3 introduces the syntax and semantics of multiple until constraint formula. Section 4 explains how to verify multiple until constraint formula based on the digraph. Section 5 gives the corresponding example description. Section 6 gives the basic algorithm description. Finally, in Section 7 we give the conclusions and directions of future research.

* Corresponding author. Tel: +86-451-82191525;
E-mail: jimingyu@hrbeu.edu.cn

2. Probabilistic reward model

In this paper, we study properties verification method of multiple until constraints formula with transition step constraint and transition reward constraint for Markov Reward Model (MRM). Below the definition of MRM is described.

Definition 1. MRM is six tuple that is divided into discrete time Markov reward model (DTMRM) and continuous time Markov reward model (CTMRM). DTMRM and CTMRM are expressed as $M=(S,P,L,AP,N,v)$ and $M=(S,R,L,AP,N,v)$ respectively, where S is a finite set of states, AP is a set of atomic propositions, $L:S\to 2^{AP}$ is a labelling function that assigns to each $s\in S$ a set $L(s)$ of atomic propositions, and $R:S\times S\to R_{\geq 0}$ is a rate matrix, $v\in Distr(S)$ is the initial distribution set and $N:S\times S\to R_{\geq 0}$ is transition reward matrix.

The transition reward of MRM is often used to represent different characteristics of the system model, such as storage space, network bandwidth, message number of successfully transmitted, power consumption, negotiation procedure and so on.

In the CTMRM, the transition probability from state s to its some successor state s' within time t meet the exponential distribution and the value is $\frac{R(s,s')}{E(s)}\left(1-e^{-E(s)t}\right)$, where $E(s)$ indicates sum of transition rate of the state s and $E(s)=\sum_{x\in S}R(s,x)$.

Because of the multiple until constraint formula studied in this paper is unrelated to the transition time, so before we verify the properties, the transition probability in CTMRM need to be pretreated, then the transition rate matrix of CTMRM is transformed into the discrete transition probability distribution of DTMRM, denoted by P_u. This pretreatment process named as discretization for continuous model.

The transformation method that transforms R of CTMRM into transition probability distribution P_u of the discrete model is divided into two cases:

(1) If $E(s')=\sum_{s''\in S}R(s',s'')\neq 0$ for some state s' in CTMRM, then in DTMRM we have $P_u(s',s'')=\frac{R(s',s'')}{E(s')}\cdot\frac{E(s')}{\max\{E(s_i)\}}$ and $P_u(s',s')=1-\sum_{s''\in S-\{s'\}}P_u(s',s'')$.

(2) If $E(s')=\sum_{s''\in S}R(s',s'')=0$ for some state s' in CTMRM, then in DTMRM we have $P(s',s'')=0$ for all states $s''\in S$.

Below the properties, quantitative validation process of DTMRM will to be shown.

3. The description of multiple until constraint formula

In the process of probabilistic model checking, the temporal logic applied in discrete probability model mainly includes PCTL and its various variants logic, in which the path formula supported transition reward and transition step is defined as:

$$\varphi ::=\Phi\left|\varphi_1\wedge\varphi_2\right|\neg\varphi\left|O_r^n\varphi\right|\varphi_1\cup_r^n\varphi_2,$$

where $n\subseteq[0,\infty]$ and $r\subseteq[0,\infty]$ respectively represent transition step interval and transition reward interval constraint, $\Delta\in\{>,<,\geq,\leq\}$ be arithmetic comparison operator, $\cup$ be until operator, O be next operator and Φ indicates the corresponding state formula.

The path formula $\varphi_1\cup_{r_1}^{n_1}\varphi_2$ is called single until constraint formula, and $\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\varphi_3$ is called double until constraint formula and $\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\ldots\cup_{r_{k-1}}^{n_{k-1}}\varphi_k\quad(k\geq 3)$ are collectively known as multiple until constraint formula.

Let σ be a path, then $\sigma\left|=\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\ldots\cup_{r_{k-1}}^{n_{k-1}}\varphi_k\right.$ established if and only if there are a series of transition processes so that: All states in the transition processes at most $Sup(n_1)$ steps satisfy the property φ_1 and the cumulative sum of transition reward during these transition steps satisfies condition r_1, and then the path migrate to one state that satisfies φ_2. On the Next transition processes, all states at most $Sup(n_2)$ steps satisfy the property φ_2 and the cumulative sum of transition reward during these transition steps satisfies condition r_2, and then the path migrate to one state that satisfies φ_3, and so on. Final, the transition processes end in the last state of σ, and the state satisfies $\varphi_k(Sup(n))$ denotes the upper bound of interval n).

As we can be seen from the semantic analysis of $\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\ldots\cup_{r_{k-1}}^{n_{k-1}}\varphi_k$ that multiple until constraint formula $\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\ldots\cup_{r_{k-1}}^{n_{k-1}}\varphi_k$ is the superposition of single until constraint formulas and the number of superposition depends on the value of k.

For the calculation of the satisfying probability of $\varphi_1\cup_{r_1}^{n_1}\varphi_2\cup_{r_2}^{n_2}\ldots\cup_{r_{k-1}}^{n_{k-1}}\varphi_k$, if path σ is starting at the state that need to be verified of the system model and can be segmented into several sub paths and each sub path

satisfies the corresponding single until formula and the sum of transition step and the sum of transition reward is within the specified constraint scope of single until formula, then the path can be named as valid path.

Then, we can calculate the transition probability of path σ, and eventually obtained the satisfy probability of multiple until constraint formula of system model.

4. The properties verification method based on directed graph

Because of the basic expression, form of the complex stochastic system model is state transition system of directed graph form, so we can verify the properties directly based on the state transition graph of the system model aiming at the characteristics of the formula to be verified.

In this paper, we consider the until formula property verification problem with the transition step constraint and transition reward constraint. In literature [6], the problem of generating counterexample of single until formula with time constraint for CTMC is described in detail and an approximation algorithm of minimal counterexample path set is given based on the structure of UDTMRM.

In ref. [7], the solving method of state satisfying probability of single until formula $\varphi \bigcup^{\le h} \psi$ with transition step constraint for DTMC is introduced in detail.

The basic idea is as follows: First, the original system model is appropriate pre-processed and is transformed into a weighted directed graph in which weight information indicates the transition probability of the original model and it is inversely proportional to the transition probability. In directed graph, the path that transition weight is least called the strongest evidence path and the process of finding the strongest evidence path is called shortest path (SP) problem. Then the solve problem of counterexample path of $\varphi \bigcup^{\le h} \psi$ formula in DTMC model is transformed into the solve problem of k shortest path set with transition step constraint for corresponding vertices based on weighted directed graph.

In this paper, the solution of satisfying probability for multiple until constraint formula $\varphi_1 \bigcup_{r_1}^{n_1} \varphi_2 \bigcup_{r_2}^{n_2} ... \bigcup_{r_{k-1}}^{n_{k-1}} \varphi_k$ will use the method similar to the above.

Definition 2. Weighted directed graph for DTMC

For DTMC $M = (S, AP, L, P)$, the weighted directed graph $D = (V, E, w)$, where vertex set $V = S$, edge set $E = \{(v, v') | v, v' \in S \wedge P(v, v') > 0\}$ and transition weight $w = -\log(v, v')$.

Definition 3. SP problem

Given a weighted directed graph $D = (V, E, w)$ and $s, t \in V$, the SP problem is to determine a path σ from s to t such that $w(\sigma) \le w(\sigma')$ for any path σ' from s to t in D. If the length of path σ and σ' satisfy $|\sigma| \le h$ and $|\sigma'| \le h$, then corresponding problem is known as the hop constrained shortest path(HCSP), which is a special case of the constrained shortest path(CSP) problem.

For the solution of SP problem, Bellman and other scholars proposed various efficient algorithms, such as Bellman-Ford algorithm and so on [8, 9]. The strongest evidence path for single until formula $\varphi \bigcup \psi$ can be solved in time $O(m + n\log(n))$ where m is the number of the edges and n is the number of the vertexes. For $\varphi \bigcup^{\le h} \psi$, the single constraint SP problem can be solved in time $O(hm)$ when $h < n - 1$.

Definition 4. K shortest path (KSP) problem

Given a weighted directed graph $D = (V, E, w)$ and $k \in R_{>0}$, the KSP problem is to find k distinct paths $\sigma_1, \sigma_2, ..., \sigma_k$ between s and t in D such that $w(\sigma_i) \le w(\sigma_j)$ for $1 \le i < j \le k$ and for every path σ between s and t, if $\sigma \notin \{\sigma_1, \sigma_2, ..., \sigma_k\}$, then $w(\sigma) \le w(\sigma_k)$.

Thus, the solution of smallest counterexample path set for formula $\varphi \bigcup \psi$ can be converted to the KSP problem and the solution of smallest counterexample path set for formula $\varphi \bigcup^{\le h} \psi$ can be converted to the HKSP problem. In literature [10], the solution algorithm of HKSP for formula $\varphi \bigcup^{\le h} \psi$ of DTMC model can be solved in time $O(hm + hk\log(m/n))$.

The properties validation method of multiple until constraint formula in this paper will use the above algorithm and has obvious differences with the existing research works in the following two aspects: On the one hand the object model be processed is different, on the other hand, the property need to be checked is different.

Therefore, we must solve the following two problems:

First, how to transform the DTMRM model for multiple until constraint formula $\varphi_1 \bigcup_{\le r_1}^{\le n_1} \varphi_2 \bigcup_{\le r_2}^{\le n_2} ... \bigcup_{\le r_{k-1}}^{\le n_{k-1}} \varphi_k$ into weighted directed graph and how to represent the transition probability and transition reward.

Second, how to describe verification algorithm of multiple until constraint formula based on the weighted directed graph.

Below, we give the detailed description of how to transform DTMRM model into weighted directed graph, which is divided into two steps:

Step 1: Adapting the DTMRM

First, we make all states in the DTMRM that not satisfy $\varphi_i (i = 1...k)$ absorbing. Then we remove all transitions that satisfy one of the following conditions:

The formal description of condition one is: $\exists k \geq m > n \geq 1\ s \in Sat(\varphi_m) \wedge s' \in Sat(\varphi_n)$ and $\neg\exists k-1 \geq q \geq 1 s \in Sat(\varphi_q) \wedge (s' \in Sat(\varphi_q) \vee s' \in Sat(\varphi_{q+1}))$.

The formal description of condition two is: $\forall k > n \geq 1\ s, s' \in Sat(\phi_k) \wedge s, s' \notin Sat(\varphi_n)$.

Next, we add an extra state t_{i-1} for each $\varphi_i (i \neq k \wedge i > 1)$ state so that each outgoing transition from φ_i state is replaced by the transition to t_{i-1} with probability 1, reward 0 and a numbered transition act and all transitions to t_{i-1} can be distinguished by the name of these transition acts. Meanwhile, we add the corresponding outgoing transitions for each $t_{i-1}(i \neq k \wedge i > 1)$ state so that each outgoing transition from t_{i-1} is transmitted to one of the states that can be transmitted from previous $\varphi_i (i \neq k \wedge i > 1)$ state with the same probability, reward and a numbered transition act that is same with one specific transition to t_{i-1}.

At last, we add an extra state t_{k-1} so that all outgoing transitions from a φ_k state are replaced by a transition to t_{k-1} with probability 1 and reward 0.

The obtained DTMRM denoted as $M' = (S', AP', L', P', N')$, where state space $S' = S \cup \{t_1, ..., t_{k-1}\}$, atomic propositions set $AP' = AP \cup \{at_{t_1}, ..., at_{t_{k-1}}\}$, labelling function $L'(t_i) = \{at_{t_i}\}$ and $L'(s) = L(s)$ for $s \in S' \wedge s \neq t_i$.

The probability matrix is divided into the following situations:

If $s \in Sat(\varphi_k)$ then $P'(s, t_{k-1}) = 1$ and $N'(s, t_{k-1}) = 0$,

If $s \notin \bigcup_{i=1}^{k} Sat(\varphi_i)$ then $P'(s, s) = 1$,

If $s \in Sat(\varphi_i) i \neq 1 \wedge i \neq k$ then $P'(s, t_{i-1}) = 1 \wedge N'(s, t_{i-1}) = 0 \wedge P'(t_{i-1}, s') = P'(s, s')$ and $N'(t_{i-1}, s') = N'(s, s')$, if $s \in Sat(\varphi_i) i \neq 1 \wedge i \neq k$, then the added orderly numbered transition acts can be denoted as $act(s, t_{i-1}) = act(t_{i-1}, s')$ when $P(s, s') > 0$.

After adjusted, the path σ that satisfies $\varphi_1 \bigcup_{\leq r_1}^{\leq n_1} \varphi_2 \bigcup_{\leq r_2}^{\leq n_2} ... \bigcup_{\leq r_{k-1}}^{\leq n_{k-1}} \varphi_k$ in original model M is turned into the path $\sigma' = \sigma_1 \cdot t_1 \cdot \sigma_2 \cdot t_2 \cdot ... \cdot \sigma_{k-1} \cdot t_{k-1}$ in the model M' and the corresponding multiple until constraint formula to be verified is turned into $\varphi_1 \bigcup_{\leq r_1}^{\leq n_1+1} at_{t_1} \bigcup_{\leq r_2}^{\leq n_2+1} at_{t_2} ... \bigcup_{\leq r_{k-1}}^{\leq n_{k-1}+1} at_{t_{k-1}}$ where $\sigma = \sigma_1 \cdot \sigma_2 \cdot ... \cdot \sigma_{k-1}$.

Step 2: Converting model into a weighted directed graph

The DTMRM model M' obtained in the first step can be transformed into a weighted directed graph by definition 3. The only difference with the definition 3 is that the weight information in the weighted directed graph of multiple until constraint formula $\varphi_1 \bigcup_{\leq r_1}^{\leq n_1} \varphi_2 \bigcup_{\leq r_2}^{\leq n_2} ... \bigcup_{\leq r_{k-1}}^{\leq n_{k-1}} \varphi_k$ not only contains the transition probability weight, but also contains the transition reward weight.

Definition 6 Weighted directed graph for DTMRM

For DTMRM $M' = (S', AP', L', P', N')$, the weighted directed graph $D = (V, E, w)$, where vertex set $V = S'$, edge set $E = \{(v, v') | v, v' \in S' \wedge P'(v, v') > 0\}$. The weight set of D denoted as $W = \{(w_p, w_N) | w_p \in W_p \wedge w_N \in W_N\}$, where transition probability weight $W_p = \{w_p | w_p = -\log(v, v')\}$ and transition reward weight $W_N = \{w_N | w_N = N'(v, v')\}$.

In the weighted directed graph of DTMRM, the computation method of transition reward weight is $w_N(\sigma) = \sum_{i=0}^{j-1} w_N(v_i, v_{i+1})$, and the computation method of transition probability weight is: $w_p(\sigma) = \sum_{i=0}^{j-1} w_p(v_i, v_{i+1}) = \sum_{i=0}^{j-1} -\log P'(v_i, v_{i+1})$

$= -\sum_{i=0}^{j-1} \log P'(v_i, v_{i+1}) = -\log P'(\sigma)$.

5. Case constructor

Consider the DTMRM which include six states and be shown in Fig. 1 and multiple until formula $a \cup b \cup c$, the model after pre-processed be depicted in Fig. 2 and the corresponding weighted directed graph be depicted in Figure 3.

The satisfaction sets of atomic propositions are $Sat(a) = \{s_0, s_1, s_4\}$, $Sat(b) = \{s_1, s_3, s_4\}$ and $Sat(c) = \{s_2, s_5\}$.

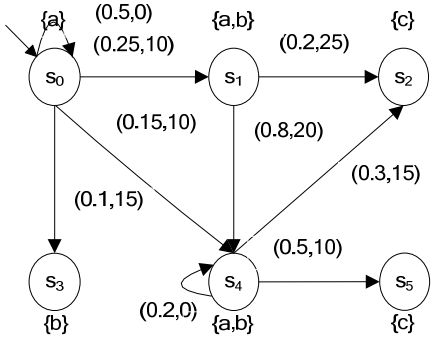


FIGURE 1. The DTMRM model to be verified

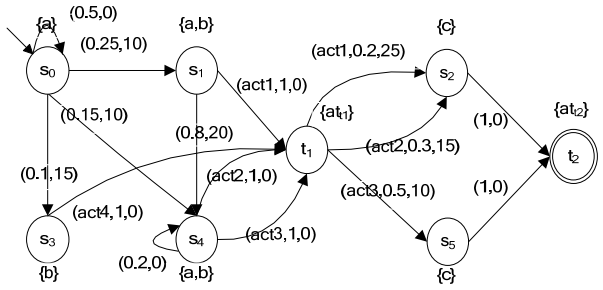


FIGURE 2. The model pre-processing for formula $a \cup b \cup c$

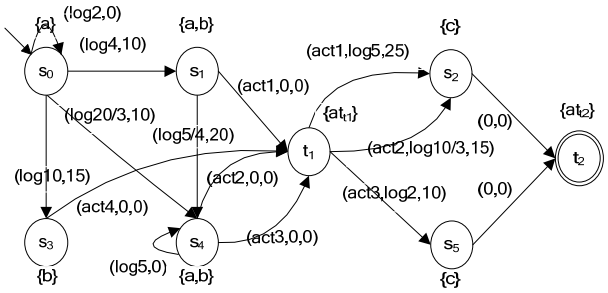


FIGURE 3. The weighted directed graph corresponding with formula $a \cup b \cup c$

6. Algorithm Description

In this paper, the satisfying path of formula $\varphi_1 \cup_{\le r_1}^{\le n_1} \varphi_2 \cup_{\le r_2}^{\le n_2} ... \cup_{\le r_{k-1}}^{\le n_{k-1}} \varphi_k$ can be considered as the connection of paths, in which each path satisfies single until formula.

Then the satisfying path set and transition probability of formula $\varphi_1 \cup_{\le r_1}^{\le n_1+1} at_{t_1} \cup_{\le r_2}^{\le n_2+1} at_{t_2} ... \cup_{\le r_{k-1}}^{\le n_{k-1}+1} at_{t_{k-1}}$ can be calculated by limiting satisfying path set of single until constraint formula with transition reward constraint on the basis of the HKSP algorithm.

Next, we roughly give satisfying path set algorithm of single until formula with transition step constraint and transition reward constraint. The algorithm named hop & reward constrained shortest paths (HRCSP), in which h denotes the upper bound of the transition step constraint, r denotes the upper bound of the transition reward constraint, ${}^i\pi_{\le h}^{\le r}(s,t)$ denotes the i-th shortest path from s to t so that the length of path is less than or equal to h and the sum of transition reward is less than or equal to r.

Algorithm HRCSP(s,t,h,r)

Require: weighted digraph D, state s and state t, and $h \in N_{\ge 0}, r \in R_{\ge 0}$

Ensure: $C = \{{}^i\pi_{\le h}^{\le r}(s,t),...\}$ with all $P = ({}^i\pi_{\le h}^{\le r}(s,t)) > 0$

1. compute ${}^i\pi_{\le h}^{\le r}(s,t)$ by BF;
2. $k = 1$;
3. $pr = P({}^i\pi_{\le h}^{k}(s,t))$;
4. **While** $((pr > 0 \;\&\&\; TRCSD(\pi_{\le h}^{k}(s,t), r)))$ **do**
5. ${}^k\pi_{\le h}^{\le r}(s,t) = \pi_{\le h}^{k}(s,t)$;
6. $k = k + 1$;
7. $\pi_{\le h}^{k}(s,t) = NextPath(t,h,k)$;
8. **End While**
9. return $\{{}^1\pi_{\le h}^{\le r}(s,t),...\}$;

In HRCSP algorithm, function TRCSD denotes transition reward constrained satisfaction decision algorithm, which is mainly used for judge the path satisfaction for transition reward constraint.

The algorithm TRCSD is described as follows:

Algorithm $TRCSD(\rho, r)$

Input: finite path ρ, reward intervals upper bound $r \in R_{\ge 0}$

Output: *true* if the cumulate reward $\rho \le r$, *false* otherwise

1. $s = first(\rho)$; $total = 0$;
2. **While** $((s' := Next(\rho, s)) != null)$ **do**
3. $totalr = totalr + N(s,s')$;
4. $s = s'$;
5. **If** $(totalr > r)$ return *false*;
6. **End while**
7. return *true*;

7. Conclusions and Future Research

In this paper, for property verification problem of multiple until constraint formulae, we put forward the construction method of weighted directed graph of MRM and the corresponding solution algorithms. The case constructor can show that the algorithms and the method are effective. In the further, we will further study the optimization problem of algorithms and extend the method in this paper to other probabilistic models.

Acknowledgements

The authors would like to thank the anonymous reviewers of this paper for their carefully reading of the manuscript as well as their many constructive comments. This paper is supported by the Fundamental Research Funds for the Central Universities (Grant No.DL11BB08), National College Students' Innovative projects (Grant No.2013 10225064); and the National Natural Science Foundation of China (Grant No. 71001023).

References

[1] Ballarini P., Mardare R,, Mura I. (2009) Analysing biochemical oscillation through probabilistic model checking. *Electronic Notes in Theoretical Computer Science*, **229**(1), 3-19

[2] Kwiatkowska M., Norman G., Parker D. (2008) Using probabilistic model checking in systems biology. *ACM SIGMETRICS Performance Evaluation Review*, **35**(4), 14-21

[3] Zhang Lijun, Jansen D. N., Nielson F., Hermanns H. (2011) Automata-based CSL model checking. *Automata, Languages and Programming*. Springer Berlin Heidelberg, 271-282

[4] Zhang Junhua, Huang Zhiqiu, Cao Zining (2008) Counterexample generation for probabilistic timed automata model checking. *Journal of Computer Research and Development*, **45**(10), 1638-45

[5] Wang Jing, Zhang Guangquan (2011) Counterexample representation for probabilistic timed automata model checking. *Journal of Soochow University (Natural Science Edition)*, **27**(2), 36-43

[6] Han T., Katoen J. P. (2007) Providing evidence of likely being on time: counterexample generation for CTMC model checking. *Automated Technology for Verification and Analysis*. Springer Berlin Heidelberg, 331- 46

[7] Han T., Katoen J. P., Berteun D. (2009) Counterexample generation in probabilistic model checking. *IEEE Transactions on Software Engineering*. **35**(2), 241-57

[8] Dijkstra E. W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1), 269-71

[9] Bellman R. (1958) On a routing problem. *Quarterly Applications Math*, **16**(1), 87-90

[10] Jiménez V. M., Marzal A. (1999) Computing the k shortest paths: a new algorithm and an experimental comparison. *Algorithm engineering*. Berlin Heidelberg: Springer, 15-29

Authors

Mingyu Ji was born in 1980, Harbin, Heilongjiang province, China

Current position, grades: Ph.D. candidate in computer science and technology in Harbin Engineering University; lecturer in Northeast Forestry University, China
University studies: He received the computer application technique master degree in 2004 from Harbin Engineering University, China
Scientific interest: specification and verification of probabilistic and stochastic systems and model checking

Yanmei Li was born in Harbin, Heilongjiang, P.R. China, 1980

Current position, grades: master's degree in computer science from Harbin Engineering University in 2008
University studies: she is now pursuing her doctor's degree in College of computer science and technology, Harbin Engineering University
Scientific interest: model checking, temporal logic

Zhiyuan Chen

Current position, grades: Ph.D. candidate in computer science and technology in Harbin Engineering University. He is currently a lecturer in the College of Computer Science and Technology, Harbin Engineering University
University studies: He received his B.S. and M.S. degrees from Jilin University in 2002 and 2005 respectively, both in computational mathematics
Scientific interest: model checking and model logic

Research on the smart wireless sensor perception system and its application based on Internet of Things

Zhang Yongheng[1*], Zhang Feng[1,2]

[1]*School of Information Engineering, Yulin University, 719000, Yulin, China*

[2]*School of Software, Xi'an University, Xi'an 710065, China*

Received 1 January 2014, www.tsi.lv

Abstract

In order to solve the uncertain perception information appears in the perception process of intelligent wireless sensor, this paper considers the intelligent perception problem of Internet of Things (IoT) based on context perception. The current status of the research on intelligent perception and its existing problem is analyzed, and then a context perception method to solve the intelligent perception problem of Internet of Thing is proposed. The intelligent perception context description model of Internet of Things is constructed. In addition, it was investigated that how the intelligent routing maintained under fault conditions, and intelligent information management system of agriculture's was proposed of agricultural IoT system, combined with agricultural automatic control devices, which had already been successfully used in the agricultural production.

Keywords: Internet of Things, intelligent perception, automatic control, context perception, intelligent wireless sensor

1. Introduction

In recent years, the Internet of Things (IoT) as an international research hotspot, have obtained broad attention. It's represents the future trend of development of the network, and requires sharing interoperability and information, so as to realize human society, the information space, the physical world ternary comprehensive connectivity and integration as the goal. Therefore, the Internet of things is regarded as the third technological revolution in information field.

Sensing technology is an important part of the field of computer science and control science, for every object to implement networking of IOT environment "can be addressed; every object network can control [1]; every spatial networking can be communication" goal, sensing technology needs formatting commands from the past context-aware simple development to the natural perception of all-round [2], three-dimensional, modes of perception from a single man-machine perception extended to man, machine, material ternary world interaction mode. Therefore, research on key technology of intelligent sensing network environment, effectively solve the problem of environment perception of things, the realization of the human society, the information space, the physical world ternary has important theoretical significance and urgent reality needs.

Over the past decade, studies carried out by China in the field of agriculture IOT technology covers the use of agricultural resources [3], agricultural ecological environment monitoring, fine management of agricultural production, agricultural product quality safety management and traceability, and other fields, but basically is at the start stage [4]. The IOT in the utilization of agricultural natural resources monitoring, based on the GPS land management and farmland information acquisition and positioning technology [1], the wireless sensor network, mobile communication network and information transmission and so on, to carry out agricultural resources network investigation of agricultural ecological environment monitoring aspect, carried out the research of soil moisture monitoring combined with ground stations and remote sensing technology based on the developed, atmospheric environment and water environment monitoring system; agricultural production of fine management [5], to carry out agricultural bio - environment information acquisition system, developed intelligent monitoring system for orchard production facilities [6]; development of livestock and poultry, aquaculture network monitoring system; developed agricultural management and traceability of the quality and safety of agricultural products [7], agricultural produce real-time information collection and transmission technology, application evaluation system and agricultural produce secure digital early warning model; electronic tag information classification and coding rules, the development of agricultural and rural consumer goods circulation regulatory information service system [8].

To sum up, agricultural IOT technology includes not only digital agriculture sensing technology, but also including wireless agricultural information network transmission

* Corresponding author. Tel: +86-135-0912-1314;
E-mail: 709863637@qq.com

technology; it is the reverse of the traditional agricultural practices, conservation of resources, and the importance of environmental protection scientific means, the main direction of the future development of agriculture. This topic using context-aware intelligent perception of key technology on the theoretical calculation of network environment, analysis of the current status of intelligent perception and the existing problems, put forward an intelligent sensing architecture based on context awareness, conducted in-depth study and on the interaction context perception, expression, fusion model and algorithm of the basic problem.

2. Intelligent Perception and Architecture of IoT Environment

Agriculture of IoT consists of three layers: that is the farmland information perception layer, information transmission layer and application layer system. The first layer is the information perception layer, including the RFID barcode, sensors and other equipment, can achieve real-time information, dynamic perception, recognition and information acquisition, perception layer mainly consists of farmland environment information collection, soil information, plant nutrition and physiological information; the second level is the information transmission layer, can realize remote wireless transmission from the Internet data information, it is mainly reflected the farmland information acquisition and transmission of large scale in the agriculture of Things [9]; the third level is the application of information system, the system can provide intelligent agriculture management by controlling the data processing and intelligent management, agricultural automation equipment, combined with the realization of intelligent agricultural production and information management, to achieve the target of save resources, protect environment and improve product quality and yield of agricultural production. Three levels of agricultural network were given an IOT to overall perception information, data transmission reliability; optimize system and intelligent information processing characteristics. Agricultural technology three levels as shown in figure 1.

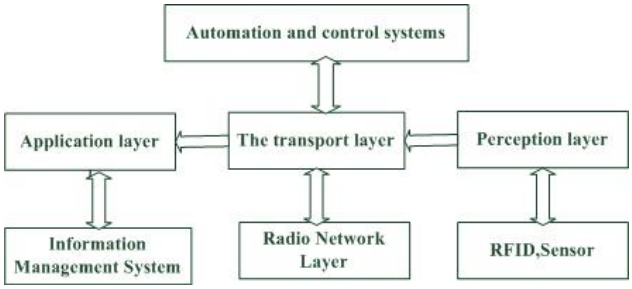


FIGURE 1 Three levels of Agricultural IOT

Intelligent perception system of IoT supports intelligent interaction the ternary of the world among of the people, machines and materials. This new type of interactive system makes the traditional IntelliSense system will face unprecedented new problems and challenges. The Internet of things to achieve the ternary world of human society, physical world, information space comprehensive connectivity and integration, it allows anyone to interaction anything at any time and any place to use any network and any service interaction.

In the IoT environment, various sensors, radio frequency identification technology, infrared sensing, global positioning system and other information generating apparatus coexist, and complex association. The user input is no longer the only trigger a system's driving force, in the IoT environment, even if the user does not sends service request, all kinds of "physical interaction" can also trigger system of intelligent computing services. In order to achieve this goal, we propose to implement the intelligent interactive system networking environment by using the context-aware technology. In this system, in addition to the user input, equipment, people, systems in various connected depending on the context interact, thus it can be seen, the networking environment context contains not only the user, system related information collection ,but also contains a collection of related information objects.

In view of the above problems, we propose the intelligent sensing architecture for Internet of things based on the context, as shown in figure 2. In this architecture, users and objects interact through the architecture of the intelligent sensing system. Where the object is divided into general devices and smart object class two, general equipment refers to the physical equipment can be identified single, one-way only provide environmental parameters for the system, intelligent objects can be perceived and physical equipment intelligent control. The technical framework of fusion rules to realize context, such as the environment changes, we only need to modify the corresponding rules without the need to modify the program code can adapt to these changes, thus increasing the intelligent interactive system adaptability.

3. Context-Aware Method of IoT

The context fusion method of IOT commonly used Bayesian estimation method and the Bayesian estimation method, neural networks and fuzzy inference, the following description these fusion methods.

3.1. BAYESIAN METHOD

The basic principle of the Bayesian approach is: given prior likelihood estimation, if we add an evidence (measurement), can be on the front (about the target attribute) likelihood estimation to be updated [10]. That is to say, with the measured value is coming, can be given assuming a priori density update for the posterior density. If $A_1, A_2, ..., A_n$ representation n exhaustive incompatible

assumptions (i.e. the presence of a target having the properties i), B is a event (or facts and observations), then the Bayesian formula is given by

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)}, \quad (1)$$

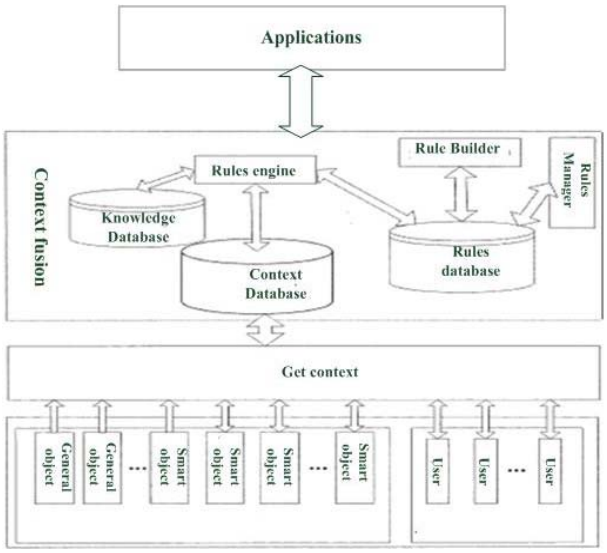


FIGURE 2. A typical electric power network

and satisfies conditions $\sum_{i=1}^{n}P(A_i)=1$,

$$\sum_{i=1}^{n}P(B|A_i)P(A_i)=\sum_{i=1}^{n}P(B,A_i)=P(B),$$

where $P(A_i)$ is the appears possibility of events $A_1,A_2,...,A_n$, and is a priori probability at assuming Ai is true condition, which is the known fact that before the experiment; $P(A_i|B)$ is given evidence of B (i target exists) conditions, assuming that A_i is true posterior probability, $P(B|A_i)$ is true condition observed probability of evidence B given A_i, $P(B)$ is the a priori distribution density of B.

Bayesian inference method to context data can be collected on the IoT Setting intelligent interactive system in multi-sensor devices and smart body fusion [11]; given assumptions to calculate the posterior probability is true. Set n sensors (they may be a different matter) work together on a target detection, set goals m attributes need to identify both of m hypotheses or propositions A_i $(i=1,2,m)$. Bayesian fusion algorithm to achieve multiple stages carried out. The first step is the collected data according to their context information obtained characteristics to identify the target attribute contact classification; the final target property is given a description of $B_1,B_2,...,B_n$. The second step is calculated for each sensor (evidence) in all assumptions true condition of the likelihood function. The third step is collecting more evidence is assumed to be true posterior probability based on Bayesian formula. The final step is the decision logic, to generate the attribute determination conclusions. The process is shown in figure 3.

In the third step, the fusion probability calculation of the target identity can be divided into two steps.

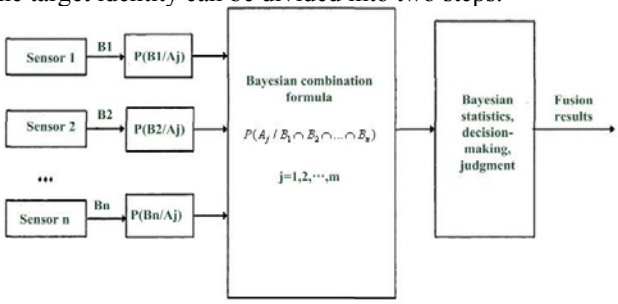


FIGURE 3. Bayes reasoning based data fusion

First, calculate the joint likelihood function of n evidence for the hypothesis under the condition of A_i, when each sensor or smart independent detection, $B_1,B_2,...,B_n$ is independent of each other, the joint likelihood function is given by:

$$P(B_1,B_2,...,B_n|A_j)=P(B_1|A_j)P(B_2|A_j)..P(B_n|A_j). \quad (2)$$

Then, using the Bayesian formula is obtained the posterior probability under the condition of n evidence, so we can write

$$P(A_j|B_1,B_2,...,B_n)=\frac{P(B_1,B_2,...,B_n|A_j)P(A_j)}{P(B_1,B_2,...,B_n)}. \quad (3)$$

The main advantages of Bayesian approach is axiomatic basis and mathematical properties of easy to understand, and only needs to calculate time medium.

3.2. NEURAL NETWORK METHOD

The neural network is to emulate the biometric information processing system for flexible information processing capability [12]. It is a microscopic simulation of the human brain function; microscopic numerical model is distributed neural network learning through a large number of samples of experience, expert knowledge and diagnosis example in the form of distribution of the weights and the closing value within the network, and the use of neural network retention to complete uncertainty reasoning. More importantly, the neural network has a strong self-learning capability, using a specific neural network learning algorithm to obtain knowledge uncertainty reasoning mechanism.

In the neural network model, radial basis function (RBF-Redial Basis Function) neural network is more representative of the meaning, it is a kind of neural network model proposed by J.Moody and C.Darken in the 80's of the last century [13], is composed of input layer, a hidden layer (radial base) and a linear output layer prior to the neural network, RBF neural network is the main feature of hidden layer radial basis function as the activation function of neurons, it has the characteristics of local experience [14], radial basis function has many forms, including the Gauss function is used more often in a radial basis function. Figure 4 shows the radial basis function neurons with n input nodes.

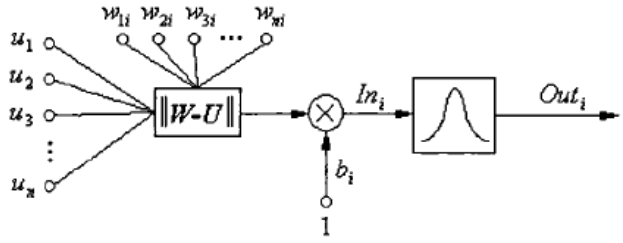


FIGURE 4. Radial basis function neurons with n input nodes

Figure 4 shows the hidden layer neurons is the vector distance between the deviation bi layer of the weight vector w and the input vector u multiplied as the input of the neuron activation function. So we can write

$$In_i = \left(\|w-u\|.b_i\right)^2 = \sqrt{\sum_{j=1}^{n}\left(w_{ji}-u_j\right)^2}.b_i \quad . \tag{4}$$

If we take the radial basis function is a Gaussian function, so we can write output neurons is

$$Out_i = e^{-\ln_i^2} = e^{-\left(\|w-u\|.b_i\right)^2} = e^{-\left(\sqrt{\sum\left(w_{ji}-u_j\right)^2}.b_i\right)^2} \quad . \tag{5}$$

By formula (4) can be seen, along with the decrease of distance between W and u, the radial basis function output value increases, and when the input is 0, which is between the W and the U distance is 0, the maximum output value is 1. Therefore, radial basis neurons as an output of one detector at the same time output of the input vector and its weight.

The radial deviation grassroots B can be used to adjust the sensitivity of the basis function, but in practical applications, more direct use is another called the stretch parameters constant two. It is used to determine each radial primary neuron to the input vector, which is the width of the radial basis function between the U and the w distance of the corresponding. Value σ (or B value) in practical applications there are many way to determine.

4. Agricultural IoT Architecture and Performance Evaluation Algorithm

4.1. AGRICULTURAL IOT ARCHITECTURE

Agriculture of Things the basic structure shown in Figure 5, the wireless sensor network architecture shown in figure 6, usually including a sensor node, a sink and a management node. Research on networking agriculture was the main star network topology.

Agricultural fertilizer and water management, the principle is similar to spraying management and automatic irrigation. Automatic irrigation system is a complex control system, their input variable is not only the water content of the soil, but there is still time, temperature, humidity, crop varieties, and crop growing season. The traditional automatic irrigation is generally the timing or timing procedures manipulation when to begin irrigation, irrigation and more often based on the experience of people for a long time may be. Soil moisture sensor control device as an automatic irrigation system, you can do the crop needs water supply, you can automatically open water, began to spray irrigation, to achieve timely sprinkler irrigation, water-saving.

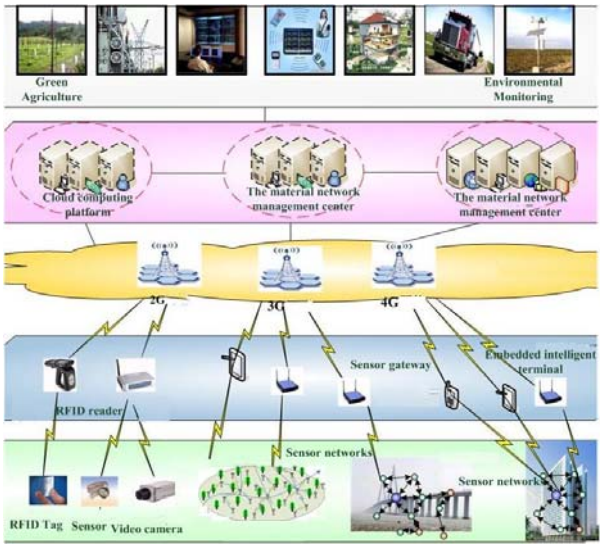


FIGURE 5. Agriculture of IOT structure

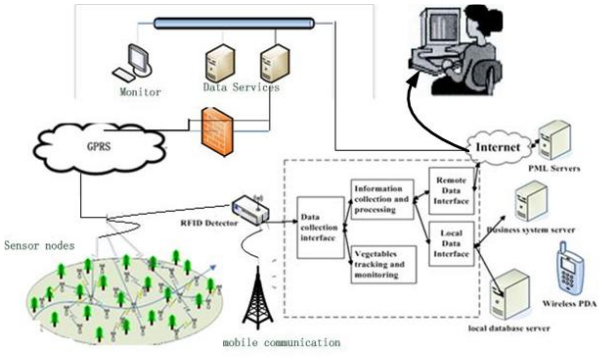


FIGURE 6. Communication basic system structure of networking

This paper focuses on the control method of automatic irrigation network information. To realize the automatic control of irrigation water, first with the real-time dynamic information acquisition field soils, in obtaining water information can be based on a variety of automatic control mode to implement the automatic irrigation. Implementation of automatic irrigation principle diagram as shown in Figure 7.

Farmland irrigation control has three control modes that are regular irrigation, automatic control and manual control. Timing irrigation is widely used, especially for the larger significance of soilless cultivation. Manual control is mainly left to the user need to manually provide manual control function, in this study, the use of manual control valve must be preset before the time limit. The diagram of information collecting node in plantation is shown in figure 8.

4.2. THE AGRICULTURE OF IOT INFORMATION COLLECTION AND TRANSMISSION OF TECHNICAL INDICATORS

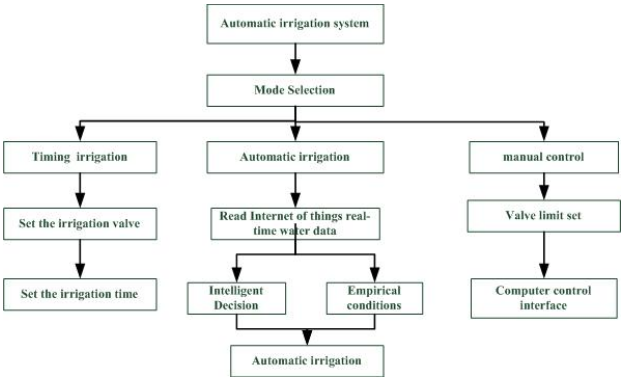


FIGURE 7. The agriculture of IOT and transmission construction

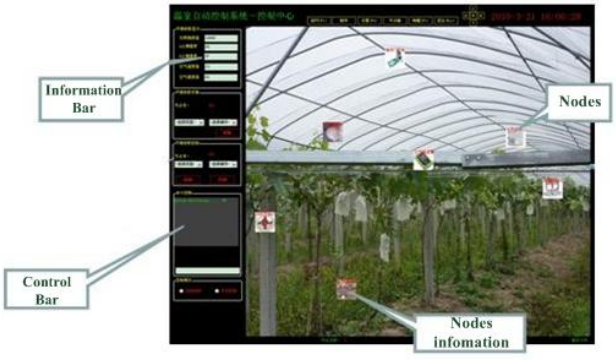


FIGURE 8. The diagram of information collecting node in plantation

Comprehensive evaluation index of IoT, according to the definition of the IOT network path, that is the assembly sequence is a $P(v_s, v_d)$ path in the G side, v_s is the source node, v_d is the destination node v_s from the source node to the destination node path v_d and is given by

$$P(v_s, v_d) = (v_s, v_1), (v_1, v_2), (v_2, v_3), \ldots, (v_{i-1}, v_i) \ldots (v_n, v_d). \quad (6)$$

Also can be a collection of nodes to represent a path simplify represented as:

$$P(v_s, v_d) = v_s, v_1, v_1, v_2, v_2, v_3, \ldots, v_i \ldots v_n, v_d. \quad (7)$$

For any one path $P(v_s, v_d)$ can define the consolidated network performance. The index contains a variety of attributes as a result of different applications, in which only consider the actual demand indicators, delay, bandwidth, and packet loss rate.

4.3. IOT NETWORK COMPREHENSIVE EVALUATION INDEX CALCULATION

IOT network performance indicators related to the average delay is $\overline{Delay}$, average bandwidth is $\overline{bandwidth}$, the average packet loss rate is $\overline{p-loss}$, the average network energy consumption evaluation is $\overline{PALL}$. However, a path of IOT road performance indicators and not fully representative of the network performance, and therefore proposed two different aspects of network comprehensive evaluation index, that is $Q_o S_{dp}$ and $Q_o S_{EB}$. Propose network-wide indicators are $Q_o s$, $Q_o s$ as a comprehensive indicator to measure the performance of the entire network situation. Where $\overline{Delay}$, $\overline{bandwidth}$, $\overline{p-loss}$ and $\overline{PALL}$ are given by

$$\overline{p-loss} = \frac{\sum_{i=1}^{m} p-loss(p(v_{si} v_{di}))}{m}, \quad (8)$$

$$\overline{Delay} = \frac{\sum_{i=1}^{m} Delay(p(v_{si} v_{di}))}{m}, \quad (9)$$

$$\overline{PALL} = \frac{\sum_{i=1}^{m} PALL(p(v_{si} v_{di}))}{m}, \quad (10)$$

$$\overline{bandwidth} = \frac{\sum_{i=1}^{m} bandwidth(p(v_{si} v_{di}))}{m}. \quad (11)$$

For a given source node and the destination node $S_d \in S$, a path $P(v_s, v_d)$ corresponding indicators following relationship:

$$Delay(p(v_s, v_d)) = \sum delay(v) + \sum Delay(e), v, e \in p(v_s, v_d), \quad (12)$$

$$Bandwidth(p(v_s, v_d)) = \min\{Bandwidth(e) + Bandwidth(v)\}, v, e \in p(v_s, v_s), \quad (13)$$

$$p - loss(p(v_s, v_d)) = 1 - \prod (1 - p - loss(e) - loss(v)), v, e \in p(v_s, v_d). \quad (14)$$

In order to specific application behind the accurately reflect and express the network performance as far as possible, the cumulative index of the network application on every path average and network tolerance compared to define comprehensive index, index of the network are often based on applications not difficult to form a fixed standard.

$$Q_o S_{DP}(p(v_s, v_d)) = \frac{\sum_{i=1}^{m} Delay(p(v_{si}, v_{di}))}{m \cdot Delay\_\max} \times \frac{\sum_{i=1}^{m} p = loss(p(v_{si}, v_{di}))}{m \cdot p\_loss\_\max}, \quad (15)$$

$$Q_o S_{EB}(p(v_s, v_d)) = \frac{\sum_{i=1}^{m} bandwidth(p(v_{si}, v_{di}))}{m \cdot bandwidth\_\max} \times \frac{\sum_{i=1}^{m} PALL(p(v_{si}, v_{di}))}{m \cdot PALL\_\max}, \quad (16)$$

$$Q_o S(P(v_s, v_d)) = Q_o S_{EB}(p(v_s, v_d)) \cdot Q_o S_{DP}(p(v_s, v_d)), \quad (17)$$

where $Delay\_\max$ specify a limit on the maximum tolerable delay for the network, $p\_loss\_\max$ specify the upper limit of the most tolerant of packet loss rate for the network specified maximum tolerated, the $bandwidth\_\max$ is the network specified maximum tolerance bandwidth usage limit. $PALL\_\max$ specify the maximum tolerable energy consumption for the network upper limit value.

4.4. IOT NETWORK PERFORMANCE TEST

Place the sensor nodes in the test of farmland large-scale farmland in 100m*600m, according to the characteristics of the actual field information collection, information acquisition and routing process the scene. The following concrete experimental plan:

In the experimental area, all the sensor nodes will transmit power amplification coefficient tune 0, signal wireless transmission distance will be significantly shortened. The test is divided into two parts, the part of the network node free networking. Test networking performance in three modes, namely: the shortest path Experimental Test (SPET), shortest path and competition for resources to optimize network the networking performance testing (O-SPET) and the depth of routing protection (D-SPET) network performance testing.

The average bandwidth occupied in three network modes experimental data as presented in Table 1, the average network delay in three network modes experimental data as presented in Table2, the average packet loss rate of the network of three nodes experimental data as presented in Table3 and the average energy consumption in three network modes experimental data as presented in Table 4.

Figure 9, 10, 11, 12 respectively is the network performance test results of three kinds of network mode. Set the $Delay\_\max = 6$, $p\_loss\_\max = 3\%$, $PALL\_\max = 30$, $bandwidth\_\max = 40$. The test results show that the shortest path (SPET) network has the worst performance, the $Q_o S = 1.21$, and the O-SPET's $Q_o S = 0.15$, compared to D-SPET's $Q_o S = 0.26$, network delay of both are similar, but the O-SPET than the D-SPET packet loss rate is low; the energy consumption is lower than that of O-SPET and D-SPET. It can be selected according to the actual situation in two different ways in different applications, routing network to improve the way. If you installed the Internet of Things adequate energy consumption on the premise, you can choose the O-SPET optimization method.

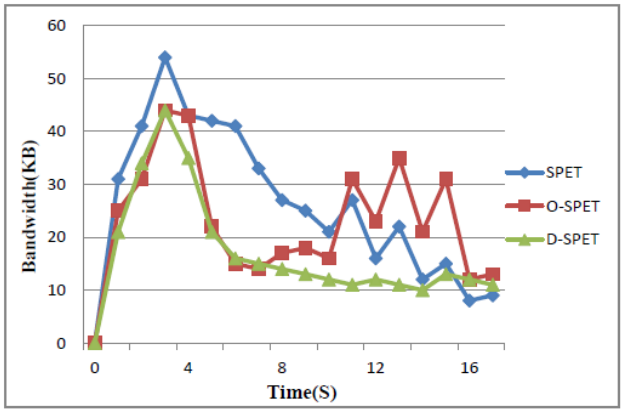


FIGURE 9. Average bandwidth occupied in three network modes

5. Conclusions

This paper proposes the use of context-sensitive technology to achieve intelligent interactive environment of IoT. IoT environment intelligent interaction throughout the entire space of human activities, any legitimate user can at any place, any time, any computing resources and services, low-cost access. People do not need to care about the specific means to achieve the calculation, they do not care about the form in which computing and services, so that humans can achieve the status of "wisdom" more refined and dynamic management of production and life, improve resource utilization and productivity levels. To presents the shortest path method, depth of prevention and resource competition law in network optimization method. Studied agricultural IOT and automatic control system combining agricultural information and automation system architecture, development the agricultural intelligent control system, intelligent management of agricultural production park automatic irrigation, fertilizer and water management, and automatic spraying; achieve the intelligent the greenhouse full-parameter intelligent control; monitoring and intelligent management of aquaculture. Some examples of successful application indicate that the intelligent control system has good application prospects.

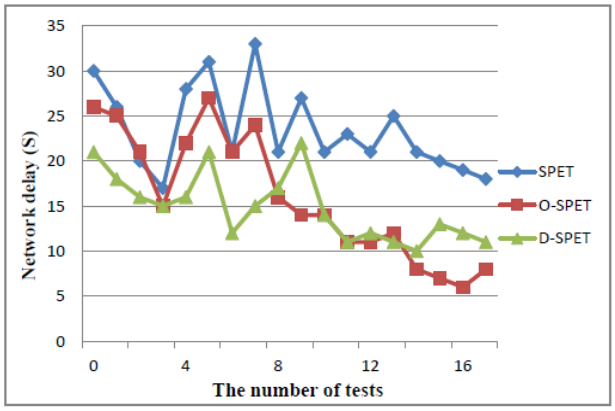


FIGURE 10. Average network delay in three network modes

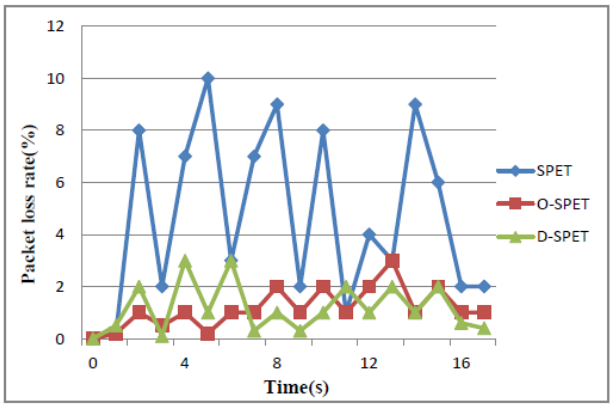


FIGURE 11. Average Packet loss rate of the network of three nodes

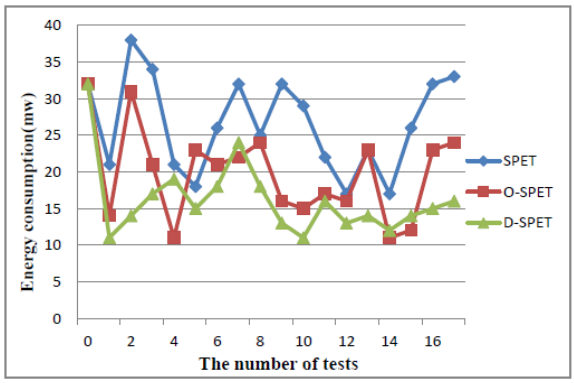


FIGURE 12 Average energy consumption in three network modes

Acknowledgments

This work is partially supported by the Funding Project for Department of Yulin University in 2013 # 13YK48, Funding Project for Department of Education of Shaanxi Province in 2013 #2013JK1167, 2012 #12JK0537, Research and Cooperation Project for Department of Yulin city (Program No. #2012cxy3-26 #Ny12-19, #Sf13-08, #Sf13-23), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2012K12-03-08). Thanks for the help.

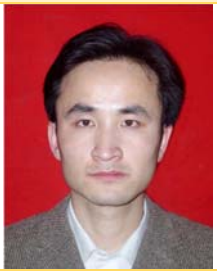
References

[1] Mahajan A., Oesch C., Padmanaban H., Utterback L, Chitikeshi S., Figueroa F. (2012) *International Journal on Smart Sensing and Intelligent Systems*, **5**(3), 559 – 75
[2] Boukhenous S. (2011) *International Journal on Smart Sensing and Intelligent Systems*, **4**(1), 21-34
[3] Zhang C. (2011) *Journal of Computers*, **6**(12), 2662-7
[4] Dakhlallah T. K., Zohdy M. A. (2011) *International Journal on Smart Sensing and Intelligent Systems*, **4**(4), 607-29
[5] Haifeng D., Xiaolong Z. (2012) *International Journal on Smart Sensing and Intelligent Systems*, **5**(4), 824-42
[6] Gao L., Zhang Y., Zhang K., Cai G., Zhang J., Shi X. (2012) A computer simulation model for Doppler ultrasound signals from pulsatile blood flow in stenosed vessels. *Computers in Biology and Medicine*, **42**(9), 906-91
[7] Huang W., Wu J. (2009) Human action recognition using recursive self-organizing map and longest common subsequence matching. In *International Workshop on Applications of Computer Vision (WACV)*, 1-6
[8] Willems G., Tuytelaars T., Gool V. (2008) An efficient dense and scale-invariant spatiotemporalinterest point detector. In *European Conference on Computer Vision (ECCV)*, **102,** 650-63
[9] Papageorgiou M., Diakaki C., Dinopoulou V., Kotsoalos A., Wang Y (2003) *Proceedings of the IEEE,* **91**(12), 2043-67
[10] Rodriguez R. I., Jia Y. (2011) *International Journal on Smart Sensing and Intelligent Systems*, **4**(2), 325-37
[11] Fathi A., Mori G. (2008) Action recognition by learning mid-level motion features. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, **2,** 726-33
[12] Yang W., Wang Y, Mori G. (2009) *In Asian Conference on Computer Vision (ACCV),* **5995,** 417- 26
[13] Abdipour A., Moradi G., Saboktakin S. (2008) Design and implementation of a noise generator. *IEEE International RF and Microwave Conference Proceedings,* December 2-4 (2008) Kuala Lumpur, Malaysia, 472-74
[14] Wetter M. (2011) *Journal of Building Performance Simulation,* **4**(3), 185-203

Authors

Zhang Yongheng, born 25.10.1968, in Shannxi Yulin

Current position, grades: associate professor in Yulin University
University studies: MS degree in Computer science from Xidian University in 2010.
Scientific interest: Data mining technology, Mass data processing technology



Zhang Feng, born 26.06.1980, in Shannxi Yulin

Current position, grades: associate professor in Yulin University
University studies: MS degree in Computer science from Xidian University in 2009. Now he is a PhD of Northwestern Polytechnical University
Scientific interest: Cloud integrated manufacturing technology, the modeling of complex systems, the Internet of Things applications

Data cleansing base on subgraph comparison

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 February 2014, www.tsi.lv

Abstract

With the quick development of the semantic web technology, RDF data explosion has become a challenging problem. Since RDF data are always from different resources, which may have overlap with each other, they could have duplicates. These duplicates may cause ambiguity and even error in reasoning. However, attentions are seldom paid to this problem. In this paper, we study the problem and give a solution, named K-radius sub graph comparison (KSC). The proposed method is based on RDF-Hierarchical Graph Model. KSC combines similar and comparison of 'context' to detect duplicate in RDF data. Experiments on publication datasets show that the proposed method is efficient in duplicate detection of RDF data. And KSC is simpler and less time-costs than other methods of graph comparison.

Keywords: RDF data cleansing, K-radius sub graph comparison

1. Introduction

The Web enables persons to link related documents. Similarly, it enables persons to link related data. Linking open data project, one of endeavors to link data, aims to use the Web to connect related data. Nowadays, according to its statistics the data sets consist of over 13.1 billion RDF triples, which are interlinked by around 142 million RDF links in 2010. Comparing with its statistics of April 2008 (2 billion RDF triples, and around 3 million RDF links) [1], the linked data have more than six times. The trends show the linked data will be even larger in the future.

However, due to diverse sources of data, it is quite possible that there exist duplicates in linked data. It may lead to many duplicate search results or wrong statistics on entities of Web etc. Let us take DBLP (http://dblp.uni-trier.de/) which data are generally integrated from IEEE, ACM Portal et al, as an example. When we search 'Jim Smith' in DBLP, DBLP returns 34 papers which DBLP thinks the single author named as 'Jim Smith' wrote. But in fact those papers were written by 5 different 'Jim Smith's. Furthermore, 'Jim Smith' is generally written as 'Smith, J.' in ACM database, but as 'J. Smith' in IEEE database. Those data make persons have to spend lots of time to explore the truth. Thus, it is highly necessary to cleanse linked data.

Linked Data uses URIs and RDF to connect pieces of data, information, and knowledge on the Web. The RDF is based upon the idea of making statements about resources (in particular Web resources) in the form of subject-predicate-object expressions. A collection of RDF statements intrinsically represents a labelled, directed multi-graph. Although some solutions were proposed to cleanse relational database, few research are found on RDF data cleansing. Graph data might be transferred to relational data, but not all graph data contain uniform and enough relationship among entities. This make the performance of traditional approaches is not very well when using those approaches to cleanse graph data.

Thus in this paper, we study the cleansing problem of linked data, and to solve the duplicate problem in linked data. The primary contributions of this article are as follows:

- Propose a simple and intuitive graph model for RDF data, named RDF-Hierarchical Graph Model, which improves Bipartite Statement-Value Graph model [5].
- Propose a simple and efficient method for graph comparison. To avoid the time-consuming comparison of graph, we introduce a new concept of K-radius sub graph. The K-radius sub graph is the 'context' of a node. We propose a K-radius Sub graph Comparison method, denoted KSC, to compare the 'context' between two nodes in RDF-Hierarchical Graph. KSC extends sub graphs of two nodes partiwisely. The sub graphs contain the relationship-information of two nodes, partiwisely. KSC compares the two sub graphs, and calculate the similarity of them. Compared to other methods, KSC is more simple and efficient for RDF graph comparison.
- Propose a solution for RDF data cleansing, based on above two. To the best of our knowledge, we are the first ones to study RDF data cleansing. In this solution, KSC, which is based on RDF-Hierarchical Graph

* Corresponding author: Tel: +86-130-988-82821; fax: +86-027-6889-3420;
E-mail: huangli82@wust.edu.cn

Model, utilizes the attribute of RDF data and the links among them, namely 'context' of a node to detect duplicate.

The rest of the paper is organized as follows. The RDF-Hierarchical Graph model is defined in section 2. Section 3 describes the proposed method K-radius Subgraph comparison method. 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.

2. Hierarchical Graph Model for RDF Data

RDF (Resource Description Framework) which is proposed by WWW Consortium* is used to describe metadata about information resource. RDF statement is a triple which is consisted of a subject, a predicate and an object. A set of RDF triples are RDF graph. RDF graph is a sort of hypergraph. In this hypergraph, the hyperedges represent the statements, and the hypernodes denote subjects, predicates and objects.

Since hypergraph is hardly analyzed, Hayes et al. proposed bipartite graphs model to represent RDF data [5]. In a bipartite graphs model, both hyperedge and hypernode are represented by a node in the graph, and they are called hyperedge-node and hypernode-node respectively. Edges in bipartite graph model are used to connect hyperedges and their hypernodes. Then the hypergraph is transformed into a bipartite graph. According to RDF, subject and object are different from predicate in types. For duplicates must be of the same type, we distinguish the two parts in a triple by layers. The nodes of same types are in the same layers. Subject and object are set on the same layer, and predicate is on a different layer. We define the layer of predicate is higher. Then we get a new model of RDF graph, called RDF-Hierarchical graph. We cannot find that there are subject of one type and object of another type in one triple. That is because subject and object of different types could not construct a statement of RDF data. Therefore, the model graph is Hierarchical.

We can take a triple as a unit in RDF-Hierarchical graph, for RDF-Hierarchical graph is consisted of triples. Figure 1(a) shows a unit in a sub graph of RDF-Hierarchical graph. Figures 1(b) (c) show different kinds of sub graphs in RDF-Hierarchical graph.

A unit is also a sub graph of RDF-Hierarchical graph. Figure. 1(b) is consisted of two units, and (c) has three units. Shadowed triples in Figure. 1(d) cannot exist.

The operations on RDF-Hierarchical graph are most unit-oriented. We will give some definitions of concepts on RDF-Hierarchical graph. Assume the target node is p, the target unit u, and G is a RDF-Hierarchical graph.

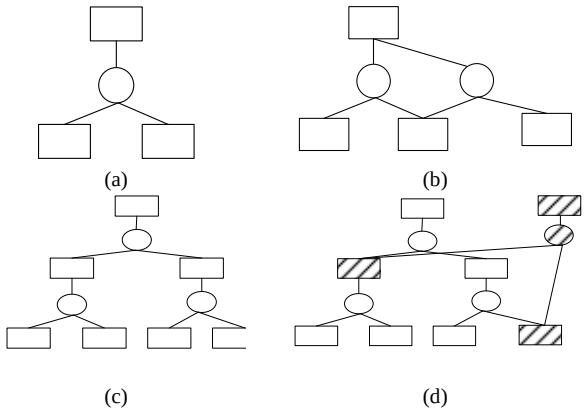


FIGURE 1. Sub graphs of RDF-Hierarchical graph model

Definition 1. n-distance. The n-distance is used to measure the distance from a unit to a node, denoted as D: $D(u, p)$

$$= \begin{cases} min\ (\{j|(u_1, u_2, ..., u_{j-1})\}) \\ \qquad p \notin u, \exists\ sequence\ (u_1, u_2, ..., u_{j-1}) \\ where\ u_i \in G, u_i \cap u_{i-1} \neq \phi, i=1,...,j\text{-}1, p \in u_1, u \cap u_{j-1} \neq \phi \\ 1 \qquad\qquad p \in u \\ 0 \qquad\qquad else \end{cases}$$

Definition 2. K-radius sub graph. The n-distance between any units in a sub graph to node P is not bigger than k. this sub graph is the k-radius sub graph of P, denoted $SG_k(P)$. If $u_i \subset SG_k(P)$, then $\max(D(u_i, P)) = k$.

In Figure 2, we give a sub graph of RDF-Hierarchical graph. From Figure. 2, we can get $D(u_1, A) = D(u_2, A) = D(u_3, A) = D(u_4, A) = 1$, and $D(u_5, A) = D(u_6, A) = 2$. We also can find all distance from units in Figure 2 to node A is not more than 2. So the sub graph shown in Figure. 2 is a 2-radius sub graph of A.

We also introduce some operations on graph. Assume $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ are two graph.

Operation 1. Subtraction. This operation, denoted as '-', means to subtract the nodes and edges in another graph. The result is a set calculated as $G_1 - G_2 = \{V_1 - V_2, E_1 - E_2\}$. The operation of subtraction is not suitable for any pairs of graphs. If $V_1 \subseteq V_2$, $V_1 - V_2$ is error. Therefore, V_1 should be larger than V_2.

Operation 2. Intersection. This operation, denoted as '$\cap$', means to intersect the nodes and edges in another graph. The result is $G_1 \cap G_2 = \{V_1 \cap V_2, E_1 \cap E_2\}$.

* http://www.w3.org/

Operation 3. Union. This operation, denoted as '$\bigcup$', means to unite the nodes and edges in another graph. The result is $G_1 \bigcup G_2 = \{V_1 \bigcup V_2, E_1 \bigcup E_2\}$.

3. K-Radius Sub Graph Comparison

Generally, in a graph, the nodes, which have smaller distance to a node P, have stronger relationship with P. These nodes restrict P, and help to disambiguate P. We call these constraints one node's 'context'. If two nodes are duplicates, they would have the same or similar 'context's. Thus, the principle of KSC is that a node is disambiguated by the 'context's, which is presented by other near nodes and edges in a graph.

The biggest difference between RDF-Hierarchical graph and a general graph is the operations are unit-oriented in RDF-Hierarchical instead of node-oriented. In RDF-Hierarchical graph, a node is disambiguated by the units around it. The reason for using units instead of nodes is that every node has inherent neighbourhoods in RDF-Hierarchical graph. The nearest nodes of them are statement nodes, which cannot help for disambiguating. According to last section, K-radius sub graph contains most units, which the n-distance from the target node is not more than k. Thus, we use K-radius sub graph to reflect the 'context' around the node.

In the following sections, we will state the method how to get the K-radius sub graph of one node, and the calculation method of similarity between K-radius sub graphs.

3.1. K-RADIUS SUBGRAPH

Here, we will describe the process of finding the K-radius sub graph of a node.

From Figure 2, we can find units are not shown clearly in RDF-Hierarchical graph. For presenting the process of create K-radius sub graph intuitively, we simplify the graph in this subsection. In the simple RDF-Hierarchical graph, each unit in RDF-Hierarchical graph is considered as a node. If two units share the same nodes, there is an edge between the two units in RDF-Hierarchical graph, and the number of nodes two units sharing is the weight of the edge. As the RDF-Hierarchical graph is hierarchy, the simple RDF-Hierarchical graph also has hierarchy. The level of the target node is level 0, and all the levels are positive. The level of a unit equals the highest level of the note in it. Figure 3 is the result simply from Figure. 2.

According to the Definition 1 and 2, the n-distance from units, which contain the target-node is 1. Therefore, these units consist of 1-radius sub graph. If k is bigger than 1, the process of finding K-radius sub graph is an iteration of extending.

However, if we try to extend all the nodes, which are within an n-distance of k to the target node, the result k-radius subgraph will be huge and complex. The goal of finding K-radius sub graph is to give some constraint to the target-node. Thus the units, which have stronger relationship with the target-node, have more worth.

In a simple RDF-Hierarchical graph, the edge expresses the two units have overlap. The weight of the edge represents the degree of the overlap. The larger the weight is the more overlap two units have. More overlap means stronger relationship. The hierarchy of RDF-Hierarchical graph also implies the degree of the relationship. The nodes in the same layer are in the same type in RDF-Hierarchical graph. The units in different layers have weaker relationship. Therefore, each time, we choose the lowest level of units. Then we conclude two properties of KSC as follows:

Assume the target node is P, $k>1$, the RDF-Hierarchical graph is G.

Property 1. The set S of candidate units is $S = \{v | \max(|v \cap u_i|), \text{ and } u_i \in SG_{k-1}(P), v \in G\}$.

Property 2. The final extending set $S' = \{t | \min(level(t)), t \in S\}$, where S is the candidate set getting from rule 1.

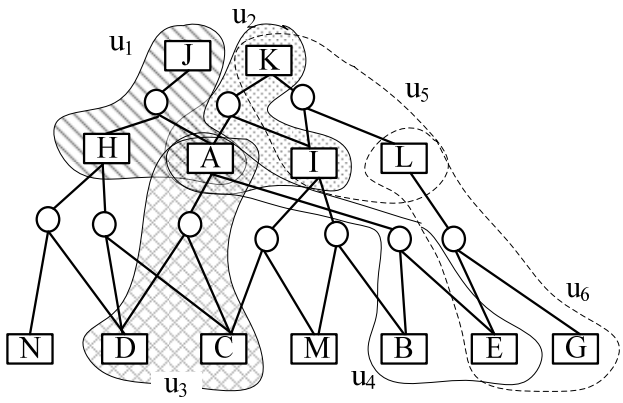


FIGURE 2. Sub graph of RDF-Hierarchical graph.

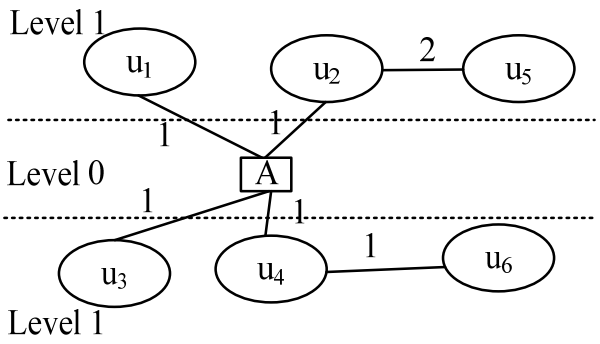


FIGURE 3. Simple RDF-Hierarchical graph

In each extending, KSC chooses the units having strongest relationship with the units, which are already in the subgraph. Therefore, in Figure 3, we will take u_6 out from the 2-radius sub graph of A.

3.2. SUB GRAPH COMPARISON

In this section, we will find the K-radius subgraph of a node, which not only shows information of a node but also helps disambiguating the node. Thus, duplicates should have similar K-radius sub graphs. According to the above analysis, if the K-radius sub graphs of two similar nodes are similar, the two nodes have a higher probability to be duplicates. Here we will give the method of comparing two K-radius sub graphs.

There are many connections among nodes in different units. These connections contain much information. Therefore, the comparison of sub graphs is node-oriented. In generally, most methods of comparing graphs are usually used in biology, and each edge refers to a regular bond. In that situation, they need only consider the construction of the graph without considering the values of nodes. That is a graph isomorphism problem. Many of them are complex and time-consuming. However, in RDF-Hierarchical graph, the values of nodes are different and very important for duplicate detection. Both construction and value should be taken into comparing of two K-radius sub graphs. More information is supplied from K-radius subgraph. It includes the nodes, the edges, and level of the nodes, the n-distance from nodes to the target node. All this information help subgraph comparison. Thus, a special method for calculating similarity between two K-radius sub graphs is needed.

Let us analyse the contributions of different information implied in K-radius sub graphs. Nodes and edges are more important for comparing. However, not all the nodes in K-radius sub graphs have the same contributions. The nearer two nodes are the stronger relationship they have. So the nodes have smaller n-distances from the target node are more important for identifying the target node. Therefore, the n-distance should be taken into calculation of similarity. The definition of n-distance orients units. The three value nodes in a unit are equal, so the nodes in a unit have the same n-distance as the unit has.

According to the above analysis, we give the calculation of similarity between two K-radius subgraph, which is shown in formula (1).

Assume two nodes are P and Q. The K-radius sub graphs are $SG_k(P)$, $SG_k(Q)$, and $SG_0(P)=SG_0(Q)=0$.

$$SimEnv\left(SG_k(P),SG_k(Q)\right)= \\ =\frac{1}{k}\sum_{i=1}^{k}\left(SimG\left(\left(SG_i(P)-SG_{i-1}(P)\right),\left(SG_i(Q)-SG_{i-1}(Q)\right)\right)\right)^i. \quad (1)$$

In formula (1), $\left(SG_i(P)-SG_{i-1}(P)\right)$ is a set, which contains the nodes with n-distances is i and the edges among them. Because SG_{k+1} contains SG_k, the subtract operation is valid on them. The intersection is to find the common nodes and edges. Along with i increasing, the nodes farther from the target-node have smaller contributions, so the Similarity increase slowly.

Since there are some ambiguity entities in the graph, it is difficult to judge that whether two other nodes in two sub graphs are the same one or not. How do we judge the two sub graphs share the node P? In [12], they proposed a method, called 'Greedy Matching' (GM), to match the nodes in two graphs. This method can avoid ambiguity problem effectively. However, that method only compares two nodes, and our comparison must consider edges as well. Therefore, we improve the GM method.

Assume two graphs $G_1(V_1,E_1)$, $G_2(V_2,E_2)$, and a threshold is ρ.

For any $v_{1i}\in V_1$, if there is a node v_{2j} in V_2, and $SimString(v_{1i},v_{2j})>\rho$, the pair (v_{1i},v_{2j}) is the candidate pair from V_1. All these pairs are cluster in set S_1. As the same, we can get the candidate set S_2 of pairs from V_2. Then according to [12], we can get the bound of $Sim(V_1,V_2)$ in formula (2):

$$USimV(V_1,V_2)=\frac{\sum_{S_1\cup S_2}SimString(v_{1i},v_{2j})}{|V_1|+|V_2|-|S_1\cup S_2|}, \\ LSimV(V_1,V_2)=\frac{\sum_{S_1\cap S_2}SimString(v_{1i},v_{2j})}{|V_1|+|V_2|-|S_1\cap S_2|}. \quad (2)$$

If two edges are similar, they could have similar nodes. Therefore, if one edge in E_1 has the similar nodes with an edge in E_2, the two edges are similar. The similarity between two edges depends on the similarities of the nodes. For any edge $(v_{1i},v_{1j})\in E_1$, we find all candidates sets of v_{1i} and v_{1j} respectively, denoted as S_i and S_j. Then put all the edges from $S_i\times S_j$, which belongs to E_2 into a set E_{ij}. Similar to the notes, we can get a bound of $Sim(E_1,E_2)$ shown in formula (3):

$$USimE(E_1,E_2)=\frac{\sum_{(v_i,v_j)\in E_1}\frac{\sum_{(v_i',v_j')\in UE_{ij}}(SimString(v_i,v_i')+SimString(v_j,v_j'))}{|US_i\times US_j|*2}}{|E_1|},$$
$$LSimE(E_1,E_2)=\frac{\sum_{(v_i,v_j)\in E_1}\frac{\sum_{(v_i',v_j')\in LE_{ij}}(SimString(v_i,v_i')+SimString(v_j,v_j'))}{|LS_i\times LS_j|*2}}{|E_1|}, \quad (3)$$

where $US_i=\{v_i' | (v_i,v_i')\in S_1\cup S_2\}$, $LS_i=\{v_i' | (v_i,v_i')\in S_1\cap S_2\}$, $UE_{ij}=US_i\times US_i$, and $LE_{ij}=LS_i\times LS_i$.

Therefore, the bound of similarity between two sub graphs SG_1 and G_2 can be calculated in formula (4):

$$USimG(G_1,G_2)=USimV(V_1,V_2)+USimE(E_1,E_2),$$
$$LSimG(G_1,G_2)=LSimV(V_1,V_2)+LSimE(E_1,E_2). \quad (4)$$

We set the average value to the similarity between two sub graphs. The calculation is show in formula (5):

$$SimG(G_1,G_2)=(USimG(G_1,G_2)+LSimG(G_1,G_2))/2. \quad (5)$$

Combining the similarity of the nodes, the final similarity of two nodes is calculated in formula (6):

$$S_{total}(P,Q)=\alpha SimString(P,Q)+\beta SimEnv(SG_k(P),SG_k(Q)), \quad (6)$$

where α and β are factors. $SimString(P,Q)$ is the similarity of strings between P and Q. If the similarity is bigger than a threshold, P and Q are duplicates.

4. Experiment

We do some experiments and show results to examine the efficiency and accuracy of the proposed approach, in this section. 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.

4.1. DATASET

We experimentally study the proposed approach on DBLP, which is a real dataset on publication. We store all data in DBLP in RDF triples. DBLP now lists more than 1.2 million publications. Since there are huge amount of triples, it is hard to measure the results by hands. Therefore, we picked up 3 groups of triples from DBLP. Each contains 5,000 triples. We use a semiautomatic method to calculate the numbers of duplications. Table 1 shows more details of the 3 groups.

TABLE 1. Detail of groups

	Number of triples	Number of entities	Number of duplicates
Group 1	5000	13057	73
Group 2	5000	2978	29
Group 3	5000	7843	57

In a RDF triple, each item is considered as an entity. In table 1, three groups contain different numbers of entities, although they contain the same number of RDF triples. This is induced by relationships among RDF triples. If all triples have fewer relationships with each other, there are fewer triples sharing entities. As we can see, the maximum number of entities in each group is 15,000 (3*5000=15000). In this situation, all the triples have no relationships with each other.

The three groups we pick up for testing represent 3 typical situations. In group 1, the number of entities is closed to the maximum number. The triples have fewer relationships with each other in this situation. In group 2, the number of entities is smaller than the number of triples. This indicates most triples share entities in this situation. The number of entities in group 3, is between the above two groups. The ratios of the number of triples to entities are shown in Figure 4.

4.2. EXPERIMENTAL ANALYSIS

The process of the proposed approach can be divided into three steps. The first step is to transfer the RDF triples into RDF-Hierarchical graph. Second, calculate the similarities of the data pairwisely. The similarity calculation has been studied in [6]. All the similarity calculation pairs are in the same layer. Here, we use the LFDW [6] to calculate similarity. If the similarity of the pair is higher than the threshold, the pair is a candidate. The candidate pairs need to be verified. The last step is to compare the 'context's between the data. The 'context' of a node is measured by the K-radius subgraph of the node. Then we combine the two similarities together as a total similarity between the two nodes. If the total similarity is bigger than the threshold, the two data are duplicates.

We first pick up two pairs of nodes in RDF-Hierarchical graph to test the similarity of K-radius sub

graphs. One pair has smaller similarity, and the other pair has bigger similarity. The details of the test data are as follows.

The test pairs: pairs 1 ("*Thomas Cormen*", "*Charles H. Leiserson*"), pairs 2 ("*John R. Smith*", "*Smith R. John*").

$$SimString(pair1) = 0.098; SimString(pair2) = 0.975$$

The result shows both similarities increase along with k in Figure. 5. This result can be easily proved by formula (1). The similarity is a summation. We also find some interesting phenomenon in the result. Similarity of K-radius sub graphs in pair 1 increases slowly at the beginning, but quickly when k is bigger than 4. And the similarity of pair 2 almost stop increasing when k is bigger than 5. After comparing strings of nodes, pair1 has less possibility to be duplicates than pair 2. When k is small, the two nodes in pair 2 share little nodes. Most units in sub graphs need to be extended. Along with increasing of K, more and more units are extended in sub graphs. The probability of sub graphs sharing nodes is higher. Meanwhile, most units of K-radius sub graphs in pair 2 stop extending. Small nodes are extended in the sub graphs. Thus, the above phenomenon happened.

From the above test, the value of k should be smaller than 4 for distinguishing two kind pairs, for the different is smaller when k is bigger than 4. Then we will test the proposed approach in different Ks, and K is smaller than 4. The results are shown in Figure 6 and Figure 7. The threshold is the same in different k.

Figure 6 shows the trend of f-measure (denoted F_1) when k increases in different groups. The trends of the f-measure are the same in three groups. The f-measures increase quickly when k is small, and slow down when k is bigger. The critical values are reached at different k in three groups. In group 1, when k is 1, the change of f-measure is small. The triples in group 1 have fewer relationships with each other. There is few information of 'context' around each entity. In this group, most duplicates are detected through similarity comparison between entity pairs. In group 3, triples share more entities with each other. Then more information of 'context' joins to help detecting. In group 3, when k is 2, f-measure reaches critical values.

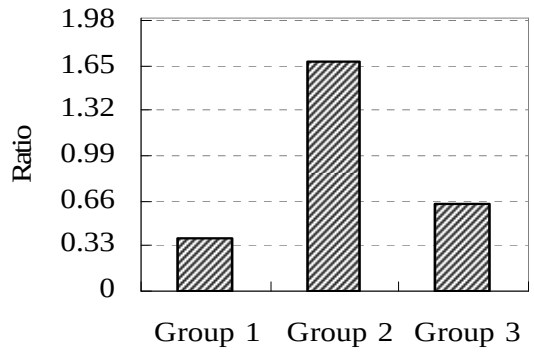


FIGURE 4. Ratios of number of 3-tuples to entities

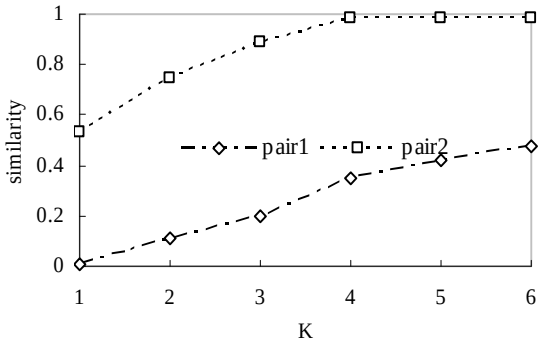


FIGURE 5. Similarities of K-radius sub graphs in two pairs

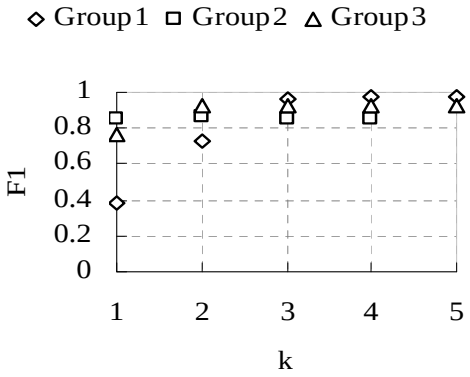


FIGURE 6. Results of detection duplicate with different k

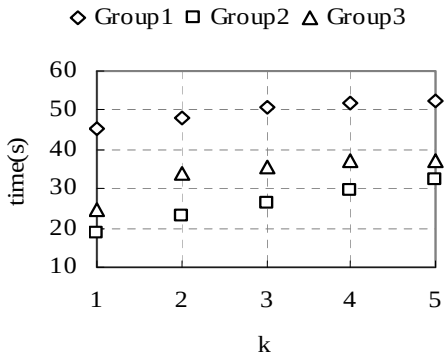


FIGURE 7. Time needed in different k

We compare the f-measures in the three groups. Group 2 is the highest, and group 1 is the lowest. This shows that 'context' is an important factor in duplicate detection.

Figure 7 shows the running times in different k. In all groups, the running time increases when k increases. The trends are similar in the three groups.

We take group 2 as an example to describe the trend in detail. The time increases slowly when k is small, quickly when k is bigger than 3, and slowly again when k is bigger than 4. When k is small, the scales of the k-radius sub graphs are small. The time cost in comparison is small. When k increases, more and more units extended to the K-radius subgraph. More time are needed for comparing. When k is still bigger, most units

in K-radius sub graphs stop extending. The scales of the K-radius sub graphs increase slowly, and then the time needed to compare is also increasing slowly. According to Figure. 6and 7, although when *k* is 4, the f-measure is highest, more time is needed. The f-measure is litter higher when *k* is 4 than *k* is 3, but much more time is needed when k is 4. Thus, the optional value of *K* for group 2 is 3.

Through similar analysis on group 1 and group 3, we can get the optional values of *k* are 1 and 2, respectively. And when the ratio of triples to entities is 1/3, there are no triples sharing entities with each other. In this situation, *k* is equal to 0. Then Figure. 8 gives the relationship between the optional value of *k* and the ratio of triples to entities.

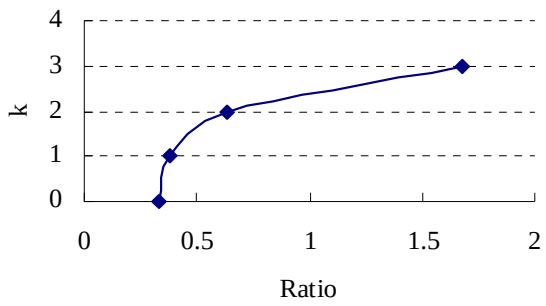


FIGURE 8. Relationship between k and ratio of triples to entities

From Figure 8, we can induce a formula to calculate *k* in a dataset. We can calculate the optional value of k using the formula (7):

$$k = \begin{cases} 0 & ratio = 1/3 \\ \lceil (\ln(ratio) + 1.1) * 1.8 \rceil & others \end{cases}, \quad (7)$$

where $ratio = \dfrac{Number_{triple}}{Number_{entity}}$.

Since RDF graph could be transfer to both Entity-relationship model and simple graph model, many methods for Entity-relationship model and graph model are also available for RDF graph. On Entity-relationship model, many methods are proposed for duplicate detecting [2, 3, 4, 9, 10] on publication data. A typical one is proposed by Han [3], which also consider the relationship among data. For graph model, Kalashnikov proposed a domain impendent method in [7], which combines the similarity and connection for identify entities. For few methods have been proposed for RDF data cleansing, we compare the proposed method KSC with methods which are used for Entity-relationship model and graph model instead. The result is shown in Figure 9. In Figure 9, for simplicity, we use '*R*' to refer to the method for entity-relationship model, and use 'C' to refer to the method proposed by Kalashnikov in [7]. '*KSC*' is the proposed approach. Here the connect-path is the shortest and least resistance path. If two units share more nodes, the connection between them is stronger. The longer the path is, the bigger resistance two nodes have, the less connection they have.

For KSC considers 'context's of data, the more 'context's the data have, the higher f-measure KSC gets. Therefore, in Figure 9(a-c), KSC on Group2 gets the highest recall, precision and f-measure. Furthermore, KSC has more advantage than other methods in Group2. The more relationships the entities have, the more effective KSC is. For there may be missing some links when transfer RDF graph to Entity-relationship model, the Entity-relationship model have lowest f-measure. The 'context' among data can better improve the precision than 'connections' do, as shown in the result.

Figure 9(d) also shows the running time of the three methods. The proposed method 'KSC' cost least time. For the number of entities is smallest in Group2, the least time is cost in Group2, although many 'context's need more time to deal with. For many extra works have to do when consider the relationship between data in Entity-relationship model, 'R' needs most time.

To present the results intuitively, we introduce another measure, called efficiency, denoted *EFF*. The calculation of EFF is shown in formula (8):

$$EFF = F_1 * Num(duplicate)/time\,. \quad (8)$$

EFF refers to the number of duplicates, which are correctly detected in a unit time. In Figure 9(e), we also show the *EFF*s of the three methods. By comparing EFFs, the proposed method 'KSC' has highest efficiency of duplicate detection in any situation. When there are more relationships among entities, the advantage of KSC is more outstanding. From Figure 9, we can find out that, the f-measure of KSC is highest, and it costs least running time in all groups. KSC is effective for any kind of dataset whether the data in it have more or less relationships.

5. Related Works

Our research solves the duplicate detection of RDF data. This research is related to two main studies: RDF data modelling and duplicate detection.

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. In [8], 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, much information would be lost. 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 [11], Morales proposed a direct 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 hard to more process on it. And in [5], Hayes proposed a bipartite graph model. This model transfers the hypergraph to a common bipartite graph. It is easy to manage and operate.

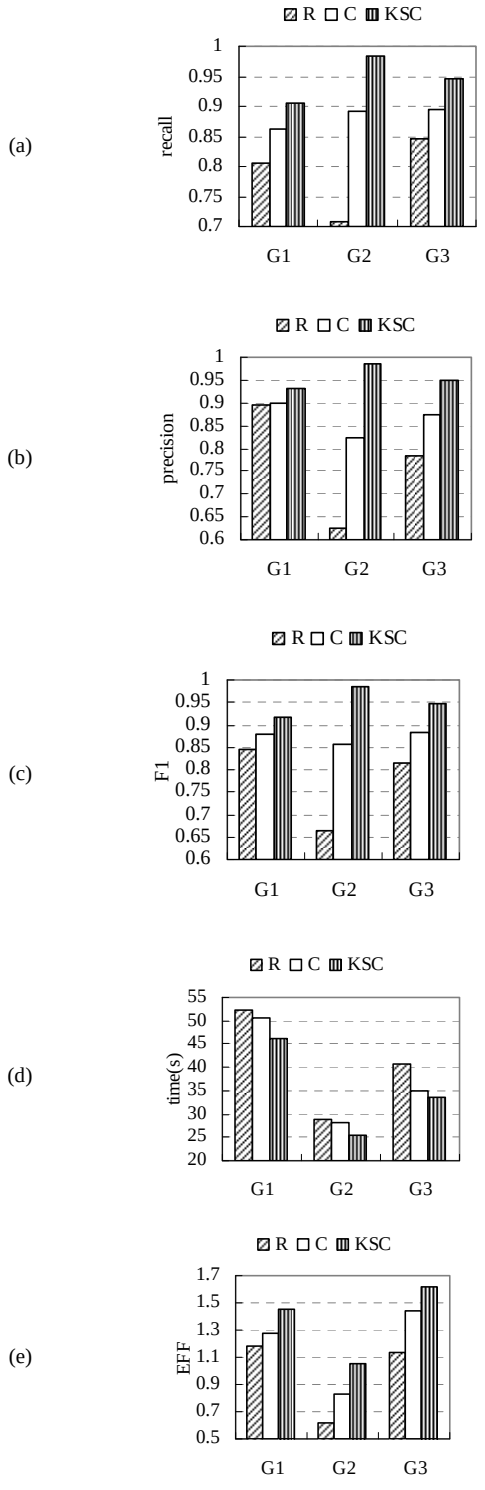


FIGURE 9. Result of comparison between R, C and KSC. Here, 'G1','G2','G3' refer to Group1, Group2, and Group3, respectively

Our model is improved from this model. All the models introduced are used for semantic retrieval, like similar query and related query. They do not focus on the duplicate detection.

Duplicate is the main inducement of data dirty. Duplicate detection belongs to 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 [9], which could be achieved by recursive field matching algorithm [10], Smith-Waterman algorithm and R-S-W algorithm [2] 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 [3]. A general object distinction methodology is introduced in [13]. 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 [4]. In [7], 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.

6. Conclusion and Future Works

Nowadays, links among data are increasing explosively. Due to variety sources of data, RDF data have duplicates. Duplicate may cause the inference and inquire error. However, studies are seldom made on this problem. Thus in this paper, we proposed an approach to detect the duplicates in RDF data.

The proposed approach combines both similarity and 'context' among RDF data to detect the duplicates. And considering the complexity of the associations among RDF data, we proposed a model for RDF, called RDF-Hierarchical graph, which is improved from Bipartite Statement-Value Graphs [4]. On the model, we give a K-radius subgraph comparison method to detect the 'context' of the two similar nodes, to avoid the complex and high cost of graph comparison. This method explores the K-radius sub graphs of the two nodes, which reflect the 'context's. By comparing two K-radius sub graphs (KSC), we get the similarity of 'context's between the nodes. Finally, we combine the similarity and the 'context' between the two nodes to decide whether they are duplicates or not.

We implement the proposed method on publication datasets, and compare the method with the methods on entity-relationship model and graph model, for seldom methods are proposed on RDF data. The results show that the proposed method improves accuracy and efficiency in detecting duplicates obviously. And the KSC is convinced to be a more simple and quick method.

Acknowledgements

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] Zander S., Schandl B. (2012) *Journal Semantic Web*, **3**(2) 131-155
[2] Franek L., Jiang X., He C. (2014) Weighted mean of a pair of clusterings. *Pattern Analysis and Applications*, **17**, 153-166
[3] Williams K., Giles C. (2013) Near duplicate detection in an academic digital library. *Proceedings of the 2013 ACM symposium on Document engineering-2013* (*Florence, Italy, September, 10-13 2013*), 334-43 http://clgiles.ist.psu.edu/pubs/DOCENG2013-near-duplicate-detection.pdf
[4] Yuan D., Mitra P., Lee Giles C. (2013) Mining and indexing graphs for supergraph search. *Proceedings of the VLDB Endowment*, 2013, **6**(10), 829-840 http://www.vldb.org/pvldb/vol6.html
[5] Gutierrez C., Hurtado C., Mendelzon A. (2011) *Journal of Computer and System Sciences*, **77**(3), 520-41
[6] Huang L., Jin H., Yuan P., Chu F. (2008) Duplicate Records Cleansing with Length Filtering and Dynamic Weighting, *The 4th International Conference on Semantics, Knowledge and Grid* 2008, Dec. 3-5, 2008, 95-102 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4725901&isnumber=4725879
[7] Ioannou E., Nejdl W., Niederée C. (2010) On-the-fly entity-aware query processing in the presence of linkage. *Proceedings of the VLDB*, **3**(1), 429-38
[8] Klyne G., Carroll J. (2004) Resource description framework (RDF): Concepts and Abstract Syntax. *W3C Recommendation. World Wide Web* http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ February, 10, 2004
[9] Minton S. N., Nanjo C., Knoblock C. A., Michalowski M., Michelson M. (2005) A Heterogeneous Field Matching Method for Record Linkage. *Proceedings of the Fifth IEEE International Conference on Data Mining (Houston Texas USA 27-30 November 2005)*, 314-21
[10] Shvaiko P. (2013) IEEE *Transactions on Knowledge and Data Engineering*, **25**(1), 158-76
[11] Morales A., Serodio M. (2007) A Directed Hypergraph Model for RDF. *Proceedings of Knowledge Web PhD Symposium* ,2007
[12] Korn F., Saha B., Srivastava D., Ying S. (2013) On Repairing Structural Problems In Semi-structured Data. *Proceedings VLDB*, **6**(9), 601-12
[13] Fan X., Wang J., Pu X. (2011) *Journal of Data and Information Quality*, **2**(2), 1242-6

Author

Huang Li, born 1982, Wuhan, China

Current position, grades: lecturer of computer science
University studies: PhD in computer science
Scientific interests: Data management, Semantic web and knowledge.
Publications : 10
Experience: 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

Numerical simulation of flow, temperature and phase fields in U71Mn rail-head quenching process

Li Gongfa*, Liu Jia, Jiang Guozhang, Kong Jianyi, Xie Liangxi, Xiao Wentao, Zhang Yikun, Cheng Fuwei

College of Machinery and Automation，Wuhan University of Science and Technology, 430081, China

Received 1 January 2014, www.tsi.lv

Abstract

With the sustainable and fast development of Chinese economy, the volume of railway freight was increasing, and which promoted the train load and speed continue improving, in order to meet the needs of high speed and over loading trains' running, heavy and seamless steel rail has increasingly widely used Heat-treatment was emphasized for its important role to qualify the products of heavy rail. And air-cooling quenching was one of the widely used heating process methods. Air-jet is a very vital instrument in quenching by air-cooling. At present, the international community has been widely used air-cooled quenching, most quenching production lines of domestic converted into wind cold quenching line. It is necessary to simulate the inner and outer flow-field of air-jet. In this paper, by means of computational fluid dynamics soft system Fluent to establish the geometric model of heavy rail, analysis the distribution of the internal and external flow field about air-jet centre cross section and the three sections were in parallel with it. Then through setting surface heat transfer coefficient of air-cooling, numerical simulation of temperature field in the cooling process of heavy rail. Finally, the phase changing temperature of steel U71Mn was got based on its CCT curves. With the cooling curves of several key points, the cooling rate at phase transition point was calculated. By comparing with every microstructure's critical cooling rate, the final cooling microstructure was predicted. Relative tests showed that the prediction was reasonable. It is significantly valuable for parameters' selection in heavy rail's technical operation.

Keywords: temperature field, flow field, phase field, air-cooling, heavy rail.

1. Introduction

As economy of our country was developing at a high speed, the railway transport was required to develop deeply. The use of heavy rail can meet the demand of large amount of load; therefore, heat-treatment of rail end was becoming more and more emphasized. Quench cooling type was one of the most important treatments in heavy rail heat-treatments, and it was also a major factor influencing quenching quality. The type of atomize-cool and air-cool are widely used in the world now. Because atomize-cool quench treatment was strictly required in controlling and sensitive to rail surface state, there was sometimes something wrong with the quality. While the velocity of air-cool treatment waves little, what's more, air-cool treatment was not sensitive to rail surface state, influenced little by artificial factor that ensure the quality of quench. The experts and scholars of domestic and foreign made many studies on quenching process. Australia's I. Elkatatny [1] by the method of computational fluid mechanics, using the software of H13 mold temperature of high pressure gas field simulation, the results obtained and the experimental results are analysed, and obtained the certain research results. The United States of America Z. Li [2] using response surface method, surface during high-pressure gas quenching heat transfer coefficient are obtained through calculation, and then for the properties of materials after high pressure gas quenching are analysed. Computational fluid dynamics method of gas quenching process of work piece was simulated, and the establishment of the three-dimensional unsteady model of vacuum high-pressure gas quenching furnace on the platform of FLUENT [3], the distribution of flow field and temperature field of work piece in furnace simulation, effects of cylindrical work piece during quenching gas type, gas pressure and speed to the cooling rate was simulated, and the simulation of quenching process [4], provides a theoretical basis for the optimization of gas quenching technology. Using the finite difference principle and experimental study of a detailed analysis of the nonlinear heat conduction problem in high pressure gas quenching steel [5-6], and then studied the synthetic surface heat transfer coefficient of the relevant issues, and consider the influence on the results of the phase transition. This paper was that by means of computational fluid dynamics soft system Fluent to establish the geometric model of heavy rail,

* Corresponding author. Tel: 18907159217; fax: 027-68862283;
E-mail: ligongfa@wust.edu.cn

simulated the distribution of the flow field, temperature field respectively on the rail air-cooled quenching, and made predict of the rail's structure, which make scientific guidance in the actual production for the rail.

2. The Establishment of Numerical Model

Air-jet device was a symmetric structure, as shown in Figure 1. 5 Air inlets with 10mm diameter, the diameter of inner hole was 0.5mm, distance was 3mm between two adjacent holes.

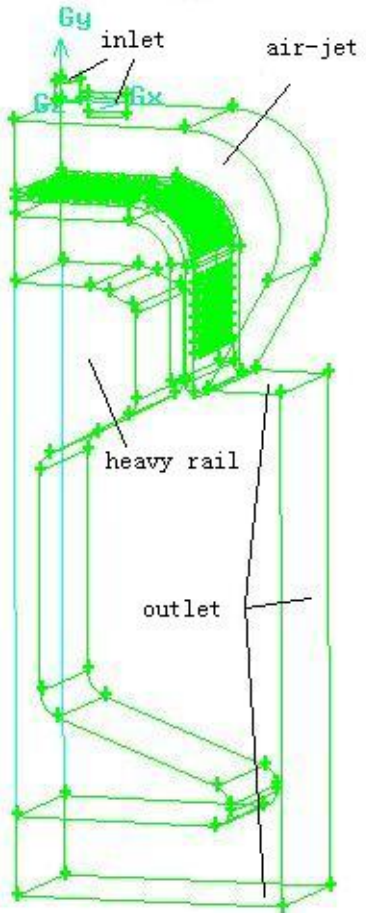


FIGURE 1. Configuration of the model

Use the Gambit that the pre-processing software of Fluent to create model. Because the model was a structure of the axial symmetry, in modelling only took 1/4 model to research. As shown in Figure 1, the lower model of air-jet was heavy rail, the heavy rail was surrounded by external flow-field, air inlet took one circle and one quarter, the air outlet was three walls. Take the an unstructured tetrahedral to mesh generation, the grid number was 2,070,000, because the diameter of small hole was too litter, the area which close to the porous panel will be block processing, in order to reduce the grid number. The wall condition of the nozzle entrance was pressure entrance, the values was pressure of gas supply, took 0.4 MPa in calculate, entrance temperature was 320 K , the export interface was pressure outlet, the pressure value was zero, and the temperature was $320K$; the interface of heavy rail and the flow field with the wall type, which use the way of convection, the heat transfer coefficient was $780W/(m^2 \cdot K)$. The reference pressure was $101325Pa$, temperature was $320K$. The researched model of this paper was 3D model, selected pressure solver, used the steady-state solver to calculate fluid, used the unsteady solver to calculate the temperature field, turbulence model took the standard k-ε turbulence model. Considering the calculation of temperature field relates to surface heat transfer, so the flow of gas to the ideal gas. U71Mn steel material density was $7800kg/(m^3)$, which Thermal conductivity of steel and the mean specific heat at constant pressure were set in the material properties with linear way respectively, the coefficient as shown in Table 2 and Table 1, the surface heat transfer coefficient of air-cooling was $780W/(m^2 \cdot K)$, and set the initial temperature of solid materials was $900^{\circ}C$ [7].

Thermal conductivity of U71Mn with the change of temperature was shown in Table 1.

TABLE 1.Thermal conductivity of U71Mn

$T(^{\circ}C)$	**100**	**200**	**300**	**400**	**500**
$\lambda(W/(m \cdot K))$	40.24	37.68	35.08	32.94	30.84
$T(^{\circ}C)$	600	700	7752	800	900
$\lambda(W/(m \cdot K))$	28.86	26	28.59	26.19	24.88

Mean specific heat at constant pressure C_P , the unit is $J/(kg \cdot ^{\circ}C)$. Mean specific heat at constant pressure of U71Mn steel as shown in Table 2.

TABLE 2. Mean specific heat at constant pressure of the U71Mn

$T(^{\circ}C)$	**100**	**200**	**300**	**400**	**500**
$C_p(J/(kg \cdot ^{\circ}C))$	478	487	505	533	579
$T(^{\circ}C)$	600	700	752	800	900
$C_p(J/(kg \cdot ^{\circ}C))$	650	772	1037	760	623

When solved the surface heat transfer coefficient of gas cooling, usually use the temperature data of work-piece that measured by experiment, and then through the finite volume difference method，nonlinear estimation method to solve. The nonlinear relationship between the surface heat transfer coefficient and surface temperature is shown in figure 2. The heat transfer coefficient was in the range of $200 \sim 1500W/(m^2 \cdot K)$. The coefficient in

$270 \sim 720^{\circ}C$ range is basically unchanged. A heat convection transfer was the main way in air-cooled [8].

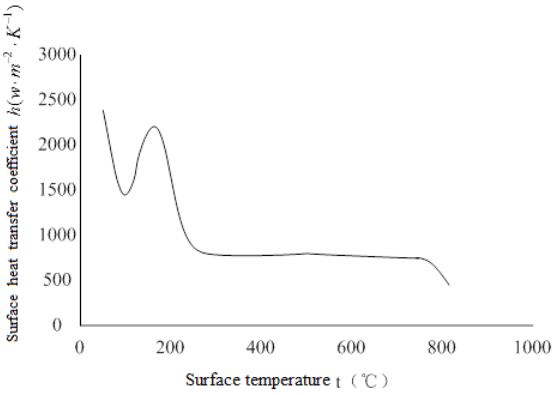


FIGURE 2. Surface heat transfer coefficient of air-cooling

As can be seen from Figure 2, for temperature from $300^{\circ}C$ to $750^{\circ}C$, a heat transfer coefficient was basically stable at around $800W/(m^2 \cdot K)$. So in this temperature range, select $800W/(m^2 \cdot K)$ for heat transfer coefficient, while at $760 \sim 810^{\circ}C$ temperature, heat transfer coefficient as a curve, selected of 4 key points numerical in the curve, in turn (745,800), (760,750), (800,500), (810,450), then, took four coordinates into the MATLAB use the method of least squares to calculate the expression of the curve. This expression was as follows:

$$y = 0.0012x^3 - 2.8x^2 + 2201x - 570505, \quad (1)$$

and, as far as data of blow $300^{\circ}C$, the software does not involve the calculation till the temperature $300^{\circ}C$, therefore not be considered. The heat transfer coefficient remains numerical value at $810^{\circ}C$ for the $810 \sim 900^{\circ}C$ just took $450W/(m^2 \cdot K)$, which the heat transfer coefficient data only test to $810^{\circ}C$ in experiment. Therefore, the expressions of heat transfer coefficient changed to this:

$$\begin{cases} y = 800, 300 \le x \le 745 \\ y = 0.0012x^3 - 2.8x^2 + 2201x - 570505, 745 \le x \le 810 \\ y = 450, 810 \le x \le 900 \end{cases}. \quad (2)$$

3. The Simulation Analysis and Results of Flow Field

The inner and outer flow field distribution at cross section Z=0 of the model and the streamlines distribution of flow field, as shown in Figures 3 and 4.

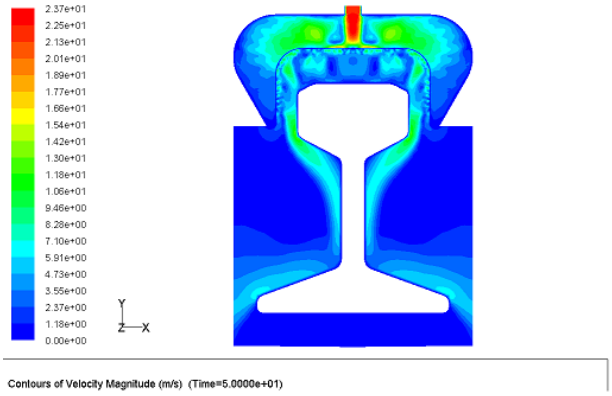


FIGURE 3. Flow field of Z=0 section

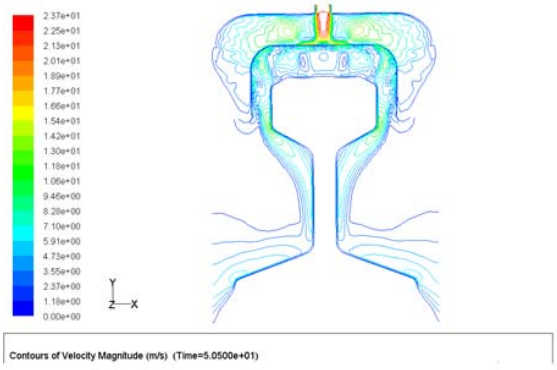


FIGURE 4. Streamlines pattern of Z=0 section

From Figures 3 and 4 can be seen，that the upper wind speed of air jet nozzle is higher，the wind speed of the holes of horizontal porous panel was much higher than it of the vertical porous panel，the speed of rail web was higher， flow field around heavy rail head was distributed uniformity.

Inner flow field and streamline pattern of X=0 section of air jet nozzle are shown in Fig.5 and Fig.6:

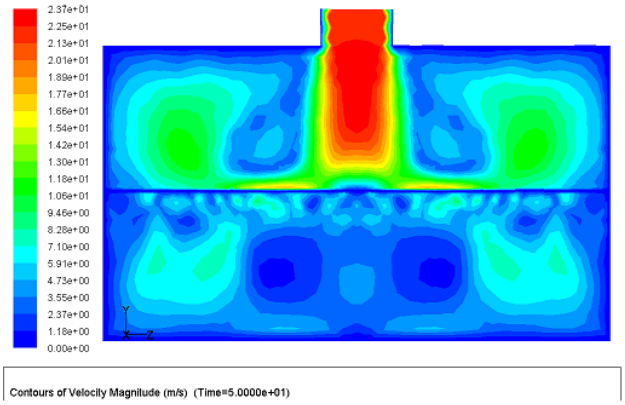


FIGURE 5. Inner flow field of X=0 section of air jet nozzle

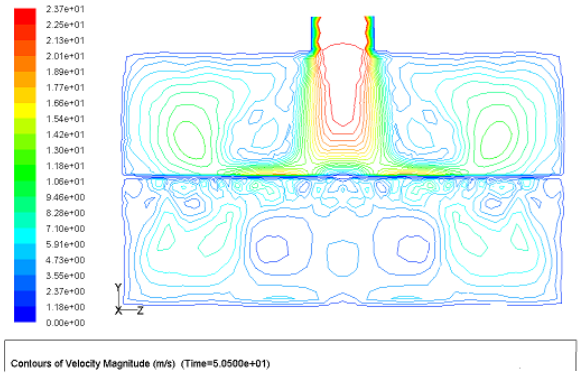


FIGURE 6. Inner streamlines pattern of X=0 section of air jet nozzle

As shown in Figures 5 and 6, when the high speed air from air inlet enters the porous plate, because the small size of the hole, most of the airflow through holes cannot flow out, but by the panel barrier to the surrounding flow, causing air flow rates of holes from the section centre 20mm significantly reduced.

The contents of the above, have studied the flow field of the centre section for a more comprehensive understanding of the flow of information, it was necessary for comparative study on flow field of each section parallel to the centre section. The following selects three sections parallel to cross section (z=0) to analysis.

The flow field of sections from the centre plane are as follows: 10mm, 20mm, 30mm.

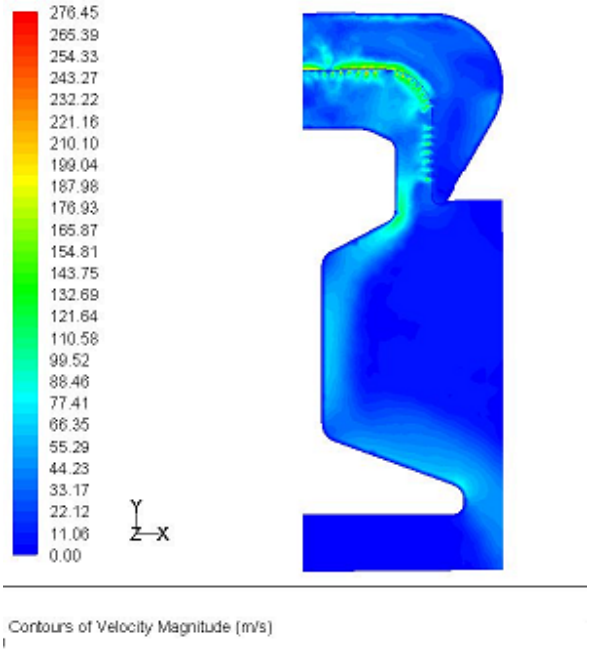


FIGURE 7. The flow field about sections of z=10mm

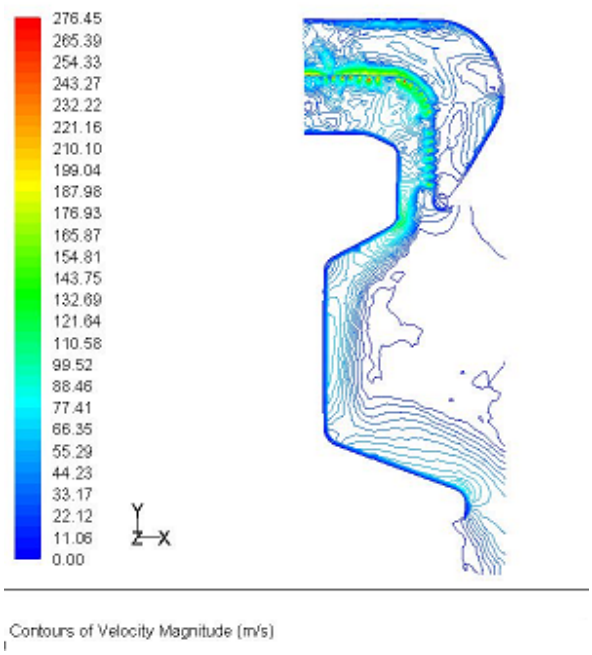


FIGURE 8. Streamlines pattern of Z=10mm

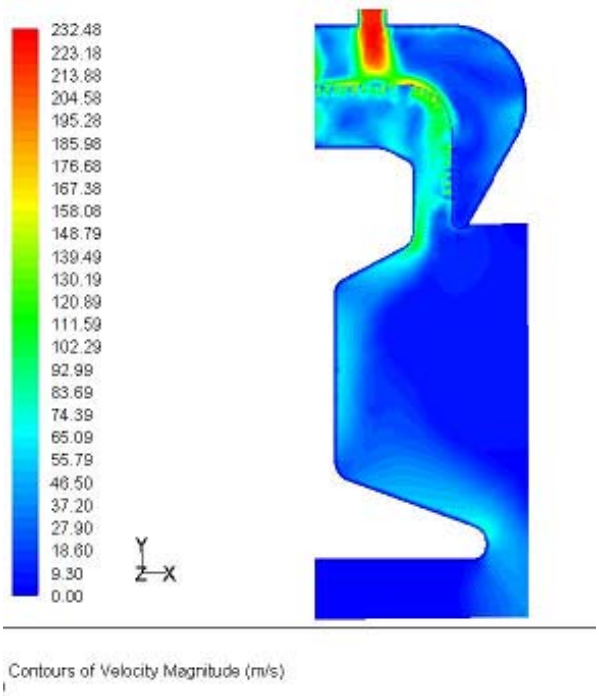


FIGURE 9. The flow field about sections of z=20mm

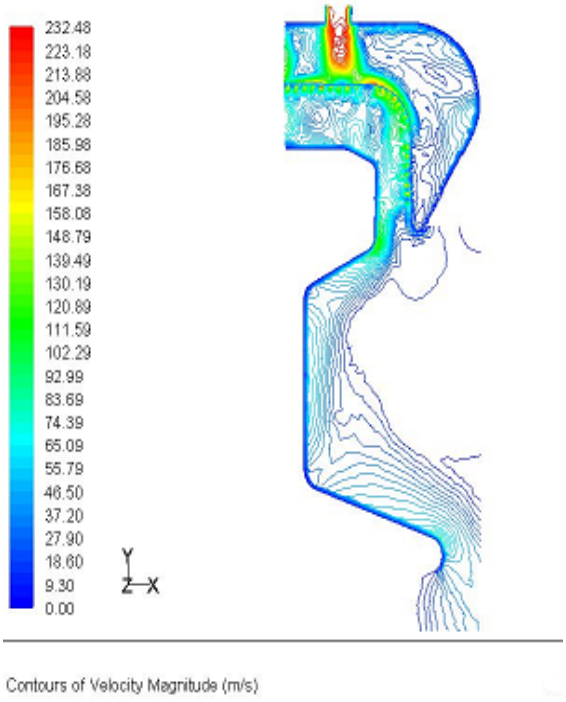


FIGURE 10. Streamlines pattern of Z=20mm

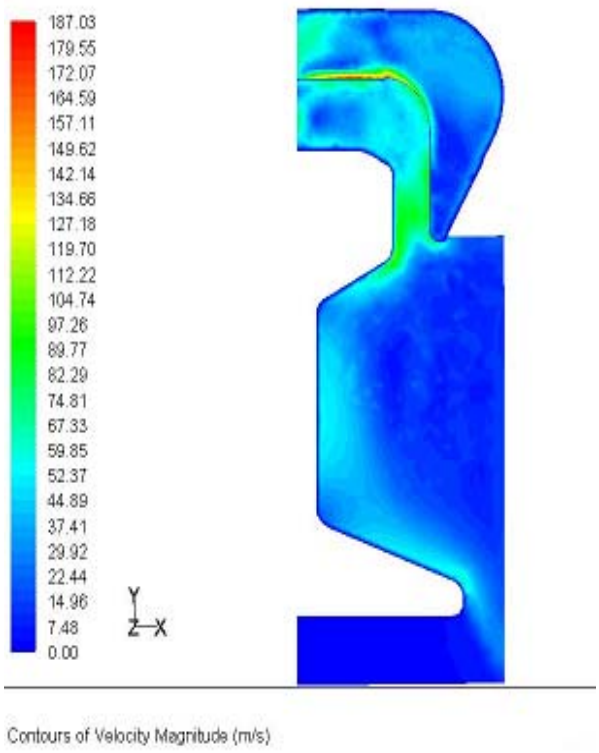


FIGURE 11. The flow field about sections of z=30mm

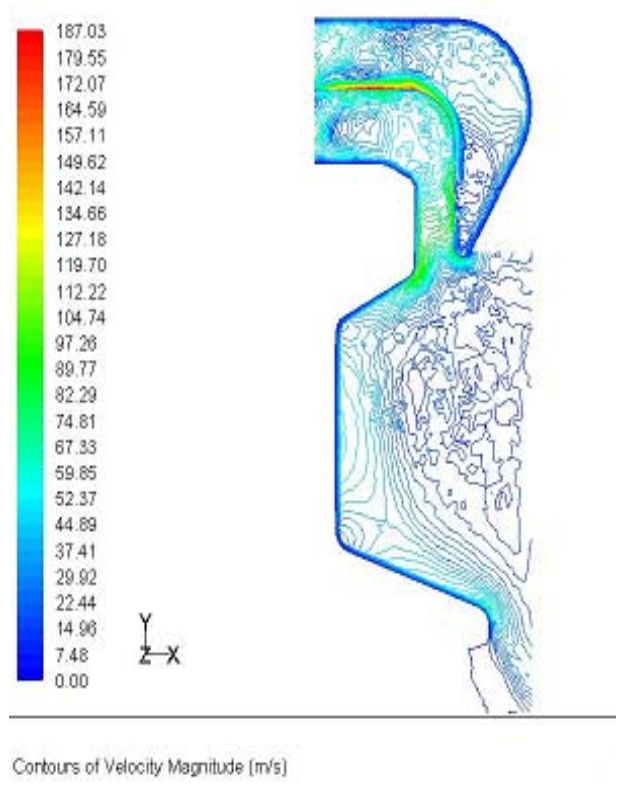


FIGURE 12. Streamlines pattern of Z=30mm

Figures 7-12 show, that the airflow become very irregular (because of the 5air inlet position). There are also differences in each cross section and longitudinal section of the flow field. Generally speaking, a flow rate near the area of air inlet was high, away from the area of the air inlet was low. Especially the flow rate of the two sharp corners area of air jet nozzle was very low. There are some rules in general about flow field near the region outside the air jet nozzle. An airflow from the porous plate in the horizontal direction directly flows to the surface of heavy rail airflow from the porous plate in the vertical direction flows to the area between the mandible and the maxilla of heavy rail in certain inclination. Then ten it flows to the rail web of heavy rail, and finally - into the rail foot [9].

4. The Simulation Analysis and Results of Temperature Field

The Figure 13 of temperature field about heavy rail was shown on the air-cooled 50s moment.

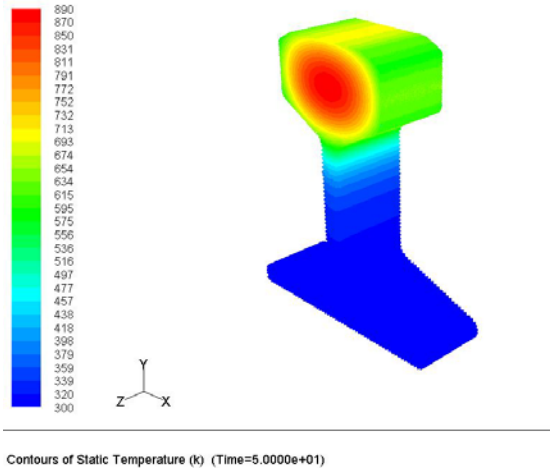


FIGURE 13. Temperature figure for heavy rail on air-cooled 50s moment

The following rules can be drawn: the part of highest temperature was located in the central of the rail head, temperature approach to 890 K that was 617$^{\circ}C$. The temperature of rail head decreases gradually from outside to inside, and the lowest temperature was 540 K , namely 267$^{\circ}C$; At the part of rail waist, the temperature reduces gradually from top to bottom, and the top temperature was 500 K , the equivalent of 227$^{\circ}C$; The foot part, the temperature was basically 300 K , temperature of 27$^{\circ}C$.

The cooling rate of some key points of heavy rail end was analysed for statistics [10], and these key points should be selected according Figure 14, setting greater distance than 10mm between two points on the tread surface and greater distance than 6mm between two points in the part of the maxilla. The selected points are marked in figure.

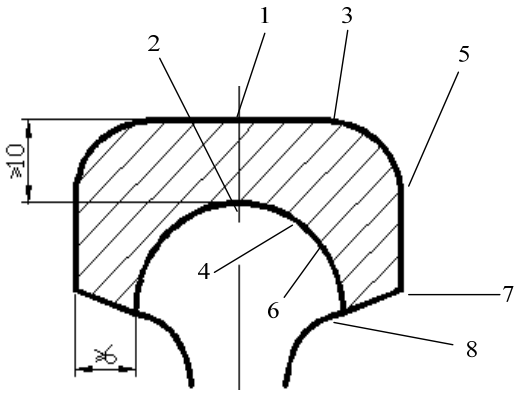
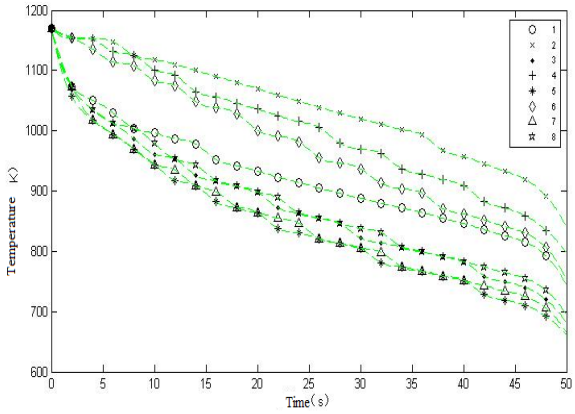


FIGURE 14. Hardening layer of rail end section and node locations



F

IGURE 15. Temperature curve of hardened layer as time

The surface temperature distribution of the whole rail accords with the test results relevant temperature field.

5. The Microstructure Field Prediction

Pearlite was formed by the community of ferrite and cementite from the eutectoid transformation of austenite. The morphology of pearlite was the layered composite material through the lamellae overlapping of ferrite slices and cementite slices, also called as lamellar pearlite. The process of air-cooling was in order to get the pearlite to enhance the comprehensive performance of heavy rail-head. The pearlite can be obtained as long as the hot heavy rail of formed austenite cools according to a certain cooling rate. The most appropriate cooling rate can not only obtain desired tissues, but also can make the internal stress to the minimum. Quench method of air-cooling can satisfy this condition [11].

Steel has three main critical cooling rates at the phase transition point:

1) The critical quench cooling rate of martensite v_k $^{\circ}C/h$: if the cooling speed was higher than v_k , the quenching steel will become the marten site;
2) The critical quench cooling rate of bainite v_3 : if the cooling speed was higher than v_3 , the quenching steel will become the bainite;
3) The critical annealing cooling rate of ferrite - pearlite v_1 : if the cooling speed was less than v_1 , the steel will become into ferrite and pearlite.

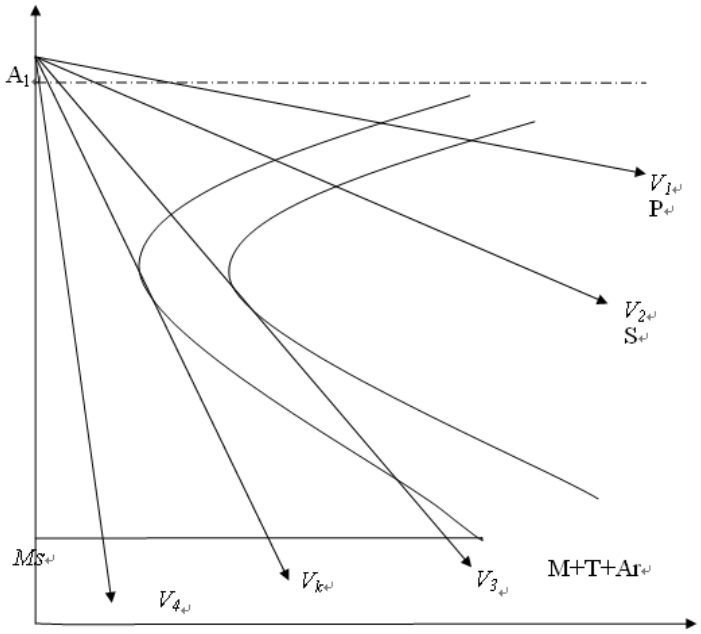


FIGURE 16. Austenite Isothermal Transformation Curve (C curve)

Studying on classification of more than 600 CCT curves was summarized by P Maynie r later on in the industry, the calculation of critical cooling rate equations was summed up:

$$\lg v_k = 9.81 - (4.62 \times C + 1.05 \times Mn + 0.54 \times Ni + 0.5 \times Cr + 0.66 \times Mo + 0.00183 \times P_A, \quad (3)$$

$$\lg v_3 = 10.17 - (3.8 \times C + 1.07 \times Mn + 0.7 \times Ni + 0.57 \times Cr + 1.58 \times Mo + 0.0032 \times P_A, \quad (4)$$

$$\lg v_1 = 6.36 - (0.43 \times C + 0.49 \times Mn + 0.78 \times Ni + 0.27 \times Cr + 0.38 \times Mo + 2.0 \times \sqrt{Mo} 0.0019 \times P_A. \quad (5)$$

Among them: Mn, Ni, Cr - the mass fraction of various alloy elements in steel body, unit 1%, P_A -- austenitising condition, unit $^\circ C/h$, numerical calculation was as follows [12]:

$$P_A = \left| \frac{1}{T_A} - \frac{nR}{\Delta H} \lg \frac{t}{to} \right|. \quad (6)$$

The critical cooling rate of the tissues was calculated, as shown in table 3.

TABLE 3. The critical cooling rate of various tissues of steel

$\lg v_3$	$\lg v_1$	$v_3(^\circ C/s)$	$v_1(^\circ C/s)$
5.03	5.74	29.85	7.16

The temperature falling curve of eight nodes was shown in the Figure 17.

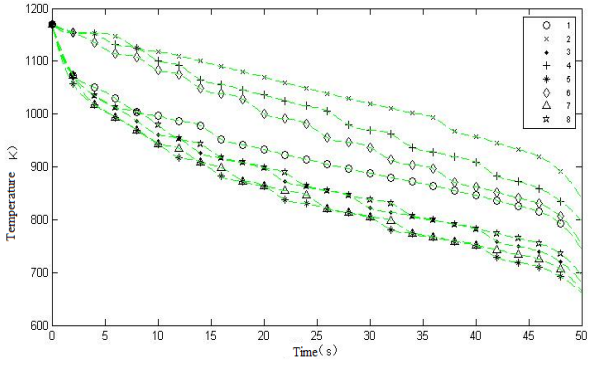


FIGURE 17. Temperature curve of hardened layer as time

According to the analysis of transition curve, phase change cooling process of the U71Mn in the beginning of the temperature was about $700^\circ C$, therefore, the three spline interpolation method was adopted to fit the points in $700^\circ C$ continuous cooling curves with MATLAB software [13]. Because the temperature unit uses the Kelvin temperature in this software, so the phase

transformation starting temperature was $973K$ and the curve equations of following points are obtained in the vicinity of the $973K$:

Node 1: $T = -x^3 + 2.6x^2 - 6.6x + 978$
Node 2: $T = -x^3 + 1.2x^2 - 5.2x + 993$
Node 3: $T = x^3 - 1.4x^2 - 10.4x + 987$
Node 4: $T = -0.7x^3 + 4.2x^2 - 5.8x + 979$
Node 5: $T = -0.1x^2 - 11.8x + 994$
Node 6: $T = -1.8x^3 + 3.3x^2 - 5.5x + 981$
Node 7: $T = -0.1x^3 - 0.8x^2 - 10x + 994$
Node 8: $T = 0.6x^3 - 4.1x^2 - 6.9x + 1004$

Based on the equation above followed by the time derivative in turn, equation of each node in the cooling rate of $700^{\circ}C$ can be obtained as following:

Node 1: $v = -3x^2 + 5.2x - 6.6$
Node 2: $v = -3x^2 + 2.4x - 5.2$
Node 3: $v = 3x^2 - 2.8x - 10.6$
Node 4: $v = -2.1x^2 + 8.4x - 5.8$
Node 5: $v = -0.2x - 11.8$
Node 6: $v = -2.4x^2 + 6.6x - 5.5$
Node 7: $v = -0.3x^2 - 1.6x - 10$
Node 8: $v = 1.8x^2 - 8.2x - 6.9$

According to the method of three spline interpolation, setting x=0, the cooling speed of each node in the $700^{\circ}C$ moment can be obtained.

TABLE 4. Nodes at $700^{\circ}C$ require the time and cooling rate

Node	The time of reach $700^{\circ}C$	The cooling rate $^{\circ}C/s$
1	15	6.6
2	38	5.2
3	9	10.6
4	30	5.8
5	8	11.8
6	25	5.5
7	8	10
8	10	6.9

According to the prediction of the quenching process model in section 4, compared the cooling rate of each node at $700^{\circ}C$ in table 4 with the critical cooling rate of each node in the table 1, quenching microstructure of each node can be predicted: for the cooling rate of region lower than $7.16^{\circ}C/s$, where nodes 1, 2, 4, 6, 8 lie, it was predicted that the quenching microstructure of these areas was fully pearlite; for the cooling rate of region between $7.16^{\circ}C/s$ and $29.85^{\circ}C/s$, where nodes 3, 5, 7 lie, it was predicted that the quenching microstructure of these areas was not fully pearlite, containing a small amount of ferrite in tissue.

6. Conclusions

In this paper, with using the Fluent software built the model of heavy rail in air-cooled quenching process, to simulate the inner and external flow field of air-jet, obtained the rule of flow field of external of air-jet, and then through setting the surface heat transfer coefficient of air-cooled, obtain the temperature field distribution of heavy rail in air-cooled quenching of 50s. That's provided certain reference value for the adjustment of parameters, which the influence of air-jet cooling. Finally, according to the CCT curves and empirical formulae, to gather statistics of the cooling rate of key nodes in hardened layer, the cooling curves were fitted with MATLAB software, and then the cooling rate at $700^{\circ}C$ was obtained by derivation and compared with critical speed of the quenching structure, then forecasted the final quenching structure. According to the related experimental results show that it was accurate prediction. The whole simulation results were accurate. Through this method to study the dimension of parameters about air-jet and predict the quenching structure, which it was better way to guide the process improvement in the production [14].

Acknowledgments

This research reported in the paper is supported by National.Natural Science Foundation of China (71271160). This support is greatly acknowledged.

References

[1] Elkatatny I. (2003) *International Journal of Thermal Sciences,* **42**(4), 417-423
[2] Li Z., Grandhi R. V., Shivpri R. (2002) *International Journal of Machine Tools and Manufacture,* **42**(5), 549-58
[3] Karlikov V. P., Sholomovich G. I. (1998) *Fluid Dynamics,* **33**(3), 313-7
[4] Bourouga B., Gilles J. (2010) *Int J Master Form,* **3,** 77-88
[5] Jingtang, Yang, Hairui, Qu ,Gongfa, Li (2010) *Hot Working Technology,* (6), 151-154 (In Chinese)
[6] Jingtang Yang, Hairui Qu, Gongfa Li (2010) *Hot Working Technology,* (3), 88-90 (In Chinese)
[7] Hua Long, Jing Tang Yang, Gongfa Li (2008) *Hot Working Technology* (22), 77-79 (In Chinese)
[8] Binkang D. U., Zhiming L. U., Pei L. U. (2013) *Journal of Mechanical & Electrical Engineering,* **30**(11), 1329-34 (In Chinese)
[9] Ying Y., Wei D., Jun L., Lan Ke (2012) *Journal of Mechanical & Electrical Engineering,* **29**(10), 1130-4 (In Chinese)
[10] Ye Yuan, Pan Wen Bin, Yang Tao, Song Jia Wang (2013) *Journal of Mechanical & Electrical Engineering,* **30**(1), 26-30 (In Chinese)
[11] Pan Wen Bin, Ye Yuan, Han Hong Jiang, Yang Tao, Song Jia Wang (2013) *Journal of Mechanical & Electrical Engineering,* **30**(1), 47-50 (In Chinese)
[12] Zhang Qi, Hou Sheng Ping (2013) *Journal of Mechanical & Electrical Engineering,* **30**(5), 600-2 (In Chinese)
[13] Kopytov E., Labendik V., Muhin V., Urbah A. (2012) Analysis of the heat processes of forming the aluminide coating on the gas turbine blades. *Proceedings of the 12th Int Conf "Reliability and Statistics in Transportation and Communication",* 17–20 October 2012, 70–77
[14] Liu J., He X., Wei X., Huang C. (2013) *Computer Modelling and New Technologies,* **17**(4), 102-11

Authors

Gongfa Li, born 10.07.1979, Honghu, China

Current position, grades: Associate professor. College of Machinery and Automation, Wuhan University of Science and Technology
Scientific interest: computer aided engineering, mechanical CAD/CAE, Modelling and optimal control of complex industrial process
Publications: 58
Experience: Dr. Gongfa Li received the Ph.D. degree in mechanical design and theory from Wuhan University of Science and Technology in China. Currently, he is an associate professor at Wuhan University of Science and Technology, China. His major research interests include modelling and optimal control of complex industrial process. He is invited as a reviewer by the editors of some international journals, such as *Environmental Engineering and Management Journal, International Journal of Engineering and Technology, International Journal of Physical Sciences, International Journal of Water Resources and Environmental Engineering,* etc. He has published nearly twenty papers in related journals

Jia Liu, born in 1990, Shanxi, China

Current position, grades: Currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology
Scientific interest: mechanical CAD/CAE, signal analysis and processing
Experience: Jia Liu was born in Shanxi province, P. R. China, in 1990. He received B.S. degree in mechanical engineering and automation from Wuchang institute of Technology, Wuhan, China, in 2012. He is currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing

Guozhang Jiang, born 12.15.1965, Tianmen, China

Current position, grades: Professor of Industrial Engineering, and the Assistant Dean of the college of machinery and automation, Wuhan University of Science and Technology
University studies: He received the B.S. degree in Chang' an University, China, in 1986, and M.S. degree in Wuhan University of Technology, China, in 1992. He received the Ph.D. degree in mechanical design and theory from Wuhan University of Science and Technology, China, in 2007.
Scientific interest: computer aided engineering, mechanical CAD/CAE and industrial engineering and management system
Publications: 100
Experience: Guozhang Jiang was born in Hubei province, P. R. China, in 1965. He is a Professor of Industrial Engineering, and the Assistant Dean of the college of machinery and automation, Wuhan University of Science and Technology. Currently, his research interests are computer aided engineering, mechanical CAD/CAE and industrial engineering and management system

Jianyi Kong, born 2.19.1961, Jiangxi, China

Current position, grades: The president of Wuhan University of Science and Technology, China
University studies: Jianyi Kong received the Ph.D. degree in mechanical design from Universität der Bundeswehr Hamburg, Germany, in 1995
Scientific interest: intelligent machine and controlled mechanism, dynamic design and fault diagnosis in electromechanical systems, mechanical CAD/CAE, intelligent design and control, etc.
Publications: 200
Experience: He was awarded as a professor of Wuhan University of Science and Technology in 1998. Currently, he is the president of Wuhan University of Science and Technology, China. He services on the editorial boards of the Chinese journal of equipment manufacturing technology. He is a director of the Chinese society for metals, etc. His research interests focus on intelligent machine and controlled mechanism, dynamic design and fault diagnosis in electromechanical systems, mechanical CAD/CAE, intelligent design and control, etc

Liangxi Xie, born 10.27.1971, Honghu, China

Current position, grades: Associate professor, College of Machinery and Automation, Wuhan University of Science and Technology
Scientific interest: rotorary vane steering gear (RVSG) and vane seals
Publications: 35
Experience: He major in mechanical design and theory and focus on the research of rotorary vane steering gear (RVSG) and vane seals. He has published more than ten papers in related journals

Xiao Wentao, born in 1989, Hubei, China

Current position, grades: Currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology
Scientific interest: mechanical CAD/CAE, signal analysis and processing
Experience: Wentao Xiao was born in Hubei province, P. R. China, in 1989. He received B.S. degree in mechanical engineering and automation from City College of Wuhan University of Science and Technology, Wuhan, China, in 2013. He is currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing

Yikun Zhang, born in 1990, Hubei China

Current position, grades: Currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology
Scientific interest: mechanical CAD/CAE, signal analysis and processing
Experience: Yikun Zhang was born in Hubei province, P. R. China, in 1990. He received B.S. degree in mechanical engineering and automation from Hu Bei University of Arts and Science, Xiangyang, China, in 2013. He is currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing

Fuwei Cheng, born 1988, Hubei, China

Current position, grades: Currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology
Scientific interest: mechanical CAD/CAE, signal analysis and processing
Experience: Fuwei Cheng was born in Hubei province, P.R, China, in 1988.He received B.S. degree in mechanical engineering and automation from Donghu college of Wuhan University, Wuhan, China, in 2012.He is currently occupied in his M.S. degree in mechanical design and theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing

An improved collision detection algorithm based on OBB

Ranran Man, Dongsheng Zhou, Qiang Zhang*

Key Laboratory of Advanced Design and Intelligent Computing

DalianUniversity, Ministry of Education, Dalian 116622, China

Received 1 January 2014, www.tsi.lv

Abstract

In this paper, we present an effective algorithm based on the Oriented Bounding Box (OBB). Particularly, the article focus on how to reduce the amount of the time that the intersection of bounding boxes are detected for increasing the efficiency of collision detection algorithm through the following three steps. Firstly, the detection time of bounding box intersection is decreased by a pre-treatment method. Secondly, we optimize the traversal approach of the bounding box tree and reduce the depth of the two fork tree under take into account the temporal and spatial correlation of motion of virtual environment. Then, we reduce the breadth of tree result from this algorithm compares the distance between the bounding boxes and traverse the node of distance that near the tree. At last, the validity of the algorithm is verified by programming simulations.

Keywords: Collision detection; Oriented bounding box (OBB); Spatio-temporal correlation; Bounding volume hierarchy

1. Introduction

Collision detection is one of the most important geometric queries, with diverse engineering applications in areas such as robotics, computer graphics, animation, computer games, virtual reality and simulation. The most simple collision detection algorithm, which is the most primitive, is a brute force method, which conduct intersection test among all basic geometric elements of two geometric models. With the complexity of the model increasing, although this method can get the correct results, the time complexity of the brute force method to O (N2) is intolerable. In recent years, domestic and foreign researchers have carried out extensive and in-depth study on the collision detection, which is already mature and widely used in space decomposition and hierarchical bounding box. The space decomposition method is the virtual space. It is divided into equal volume of small units, only to test the geometric objects occupying at the same unit and adjacent cell test. Typical examples are the uniform grid, k-d tree, octree, BSP tree. The space decomposition method is usually applied to the collision environment where geometry objects are sparse. Hierarchical bounding box method is a widely used method in collision detection. In order to nearly completely to obtain the object's geometry characteristic, the basic idea is to use slightly larger simple bounding box whose geometric feature volume approximation to describe the complex geometric objects, via a hierarchical tree to approach the geometric model. So only, the parts of bounding box overlap need further test.

There are some typical methods; such as AABB oriented bounding box [1], sphere [2], OBB and k-DOPs [3]. The OBB bounding box was widely used, for the advantages of OBB that whose compact package and updating required only a simple shift operation. This paper mainly researched about the OBB bounding box, and put forward some improvement methods. In recent years, researching of OBB based on bounding box mainly summarized as follows.

With the development of computer graphics, to solve the problem of collision detection is widely used by the parallel GPU technology [4-6, 16]. In addition, Zhang X [17] uses graphics hardware technology for the collision detection algorithms. These methods seem to pose an interesting alternative to object – space methods. However, discretization errors can cause such algorithms report an incorrect subset of all collisions. Each bounding box method has their pros and cons, taking the advantages and disadvantages of the bounding box, which can foster strengths and avoid weaknesses. For example Rui Huang [7], Chaoqiang Tu [8] combine AABB and OBB bounding box, achieve good results. Jung-Woo Chang [9] and Qian-Ru Xie [10] combine with OBB and spherical bounding box. In addition, OBB collision detection includes triangle intersection problem. To improve the detection of triangle intersection will also accelerate the speed of algorithm. To speed up the algorithm, Jung-Woo Chang [11] takes the problem of triangle intersection as the main research object. There are also some scholars applying the genetic algorithm to the field of collision detection, such as PSO algorithm, to accelerate OBB collision detection [12, 13]. Ying-Mei W [14] first takes advantage of the temporal correlation to the field of collision detection. Then some people such as Guo Lingyun [15] use space-time correlation to

* Corresponding author: Tel. 8641187403733,
Email: zhangq30@yahoo.com

accelerate the Collision detection of multiple moving objects, and have achieved certain results.

Collision detection algorithm based on bounding box is an important algorithm of collision detection. This paper focuses on the study of the flow of OBB bounding box algorithm, analyse how to improve the OBB bounding box algorithm, and takes advantage of temporal and spatial correlation to improve the OBB algorithm, and at last use experiment to verify.

2. Improving Algorithm Based on OBB Hierarchical Bounding Box

Oriented bounding box is first used in 1996 to achieve the "RAPID" system by Gottschalk. It was first applied in ray tracing, and can be closely attached around the closed object. Collision detection based on OBB bounding box is one of the few and fast detection system.

2.1. OBB BOUNDING BOX ALGORITHM FLOW CHART

The bounding box surrounding objects method calculate bounding volume before the test based on geometric primitives, which will effectively improve the test performance. However, the test process is simplified; the number of tests did not change between the two and two groups. At the same time, the test results remain asymptotic time complexity of the original. Bounding box technology only adjusts the test process in a constant level. Hierarchical bounding box technology was integrated into a tree structure. In this way, the time complexity can be reduced to log level [18, 19].

The algorithm based on the flow diagram of the OBB hierarchical bounding box tree is shown in Figure1.

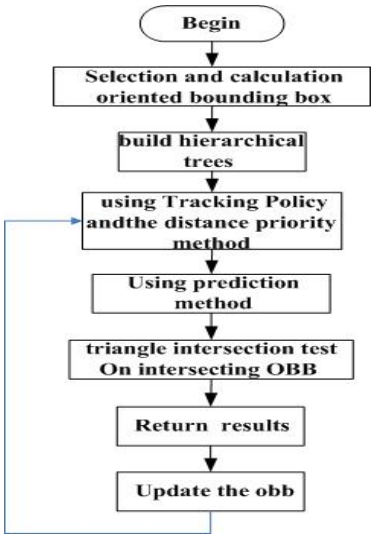


FIGURE 1. Program flow char of OBB collision detection

Take it as the pre-processing phase of collision detection that calculating the OBB bounding box and creating a hierarchical bounding box tree. Take the process of detecting collision as the phase of collision detection. The key of OBB calculation is to find the optimal direction, determine the minimum size of bounding box enclosing the object in the right direction. Establish hierarchical bounding tree is to establish a hierarchical of collision detection. Usually the two binary tree traversal method was used to product oriented bounding box trees.

2.2. COLLISION TESTING USING HIERARCHY TRAVERSAL

Given the environment hierarchy, bounding volume tree E, and the flying hierarchy, bounding volume tree F (after tumbling), we must traverse the two trees efficiently to determine if any part of the flying object collides with some part of the environment. The algorithm is outlined in Algorithm 1. It consists of a recursive call to TraverseTrees (vF, vE), where vF is the current node of the flying hierarchy and vE is the current node of the environment hierarchy. Initially, we set vF and vE to be the root nodes of the hierarchies.

At a general stage of the traversal algorithm, we test for overlap between the bounding volume b(vF) and the bounding volume b(vE). If they are disjoint, then we are done with this so as to call to the function. Otherwise, if vE is not a leaf, we step down one level in the environment hierarchy, recursively calling TraverseTrees(vF,vE) for each of the children ve of vE. If vE is a leaf, then we check if vF is a leaf: if it is, we do triangle-triangle intersection tests between each triangle of vE and each triangle of vF; otherwise, we step down one level in the flying hierarchy, recursively calling TraverseTrees(vf,vE) for each of the children vf of vF .

Algorithm TraverseTrees(vF,vE):

Input: A node vF of the flying hierarchy, a node vE of the environment hierarchy

(1) if b(vF)∩b(vE) ≠ϕ then
(2) if vE is a leaf then
(3) if vF is a leaf then
(4) check test triangles tE and tF for intersection
(5) for each child vf of vF
(6) TraverseTrees(vf,vEE)
(7) else
(8) for each child ve of vE
(9) TraverseTrees(vF,ve)
(10) return

We can see that, there are bounding box intersection tests and triangle intersection tests in the process of traversing the OBB bounding box tree. The bounding box intersection test is based on the separation axis theory. If an OBB has a separating axis, you can determine which of the two OBB are not overlapped. For a pair of OBB, there are 15 tests to be separated (3 surfaces of each OBB plus the 3 sides of the two OBB combinations). Finding such, a separating axis can judge that OBB is not overlapped. If the OBB overlap, it required a 15 test. During the process of OBB bounding box tree traversal,

large numbers of intersection tests based on bounding box are need. The basic triangle intersection test was conducted when the cycle into the leaf node. Through the analysis of the OBB hierarchy bounding box algorithm, there are large number of bounding boxes in hierarchical tree. If you can reduce the OBB bounding box test number, the efficiency of the algorithm will improve. The improved algorithm is proposed based on the following.

2.3. PRETREATMENT WITH BOUNDING BOX INTERSECTION TEST

Intersection test bounding box need 15 times comparison operations, 60 times addition and subtraction, 81 times multiplication and 24 times the absolute value operation. So add a pre-processing method of two OBB bounding boxes of intersection. OBB bounding box technology can well wrap the object according to the shape, compared with other bounding box. However, from another side, the storage of OBB bounding box is large than others, the complexity of the bounding box intersection test is the highest.

The sphere bounding box is another bounding box technology, it can effectively compensate for the shortcomings of OBB bounding box. Although the sphere bounding box tightness is relatively low, the structure of sphere bounding box storage structure is less, the intersection test is fast, and updating the sphere bounding box is not influenced by the rotation transformation, the sphere bounding box is often used for collision detection. For the advantage of sphere bounding box intersection test, the OBB bounding box based on the ball, using a rough prediction bounding box intersection test of the ball, which can avoid complicated calculation of bounding box intersection test, can improve the efficiency of the algorithm. Specific ideas are as follows:

Take bounding box centre point as the centre; get the ball radius (r1, r2) through the calculation between the farthest vertex and centre distance. Similarly take the shortest distance of the bounding box centre as the radius, take the centre as the centre, and draw a ball (rmin1, rmin2). When the distance between the two bounding box (re) is greater than the distance between spheres, bounding box does not intersect. When the bounding box is less than the distance between the balls, the distance between the bounding boxes will intersect. When the distance between the bounding boxes is located between the two, use the method of separating axis. Algorithm of two-dimensional diagram is shown in Figure 2, in which two rectangles represent two bounding box, embedded and wrapped in rectangular representing the big ball and small ball.

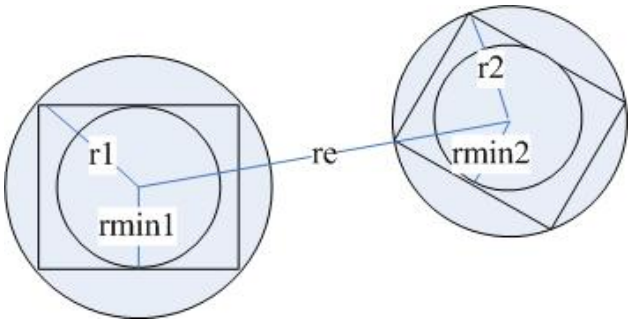


FIGURE 2. 2D sketch map of two OBB bounding box position

The main algorithm using spherical bounding box is shown as follows:

If (re<rmin1+rmin2) / / when the bounding box is less than the distance between the balls, return bounding boxes intersect.

{return 0.}

Else if (re>r1+r2) / / when the bounding box is greater than the distance between the balls, return the bounding boxes do not intersect

{return 1;}

else / / or call the algorithm of separation axis intersection test

{return obb_disjoint (R, T, b1->d, D);}

The formula of calculating of the two bounding box distance re is shown in Equation (1). Because the bounding box is in different coordinate system, the two bounding box should be converted into the same coordinate system. In order to ensure the correctness and robustness of the algorithm, add some fill factor in the calculation of re for the reason of the algorithm computing the square root.

$$re = \sqrt{(x1 = x2)^2 + (y1 - y2)^2 + (z1 - z2)^2} + \varepsilon . \quad (1)$$

The point (x1,y1,z1) and (x2,y2,z2) were pointed to the origin of coordinate of the two balls into the same coordinates. ε makes up the factor.

The formula of converting origin of two spherical coordinates to a coordinate is shown in Equation (2).

$$P_{(2)}R_{(2->1)} + T_{(2->1)} = P_{(1)} . \quad (2)$$

The $P_{(2)}$ is called the second bounding box origin coordinates. $R_{(2->1)}$ and $T_{(2->1)}$ point to the rotation and translation matrix from the first coordinate to second coordinate system respectively. $P_{(1)}$ represents the point that the second coordinate system $P_{(2)}$ point conversing to the first point in the coordinate system.

2.4. BY SELECTING THE NEAREST NODE TRAVERSAL,REDUCE THE WIDTH OF TREE TRAVERSAL STRATEGY

As is mentioned before, in the process of bounding box traversal, we stop traverse depth bounding box tree when a node is disjoint. If join optimization traversal strategy in the process of traversing bounding box tree, so as to traversal the bounding box tree in accordance with the most likely path of collision, which can accelerate the traversal speed.

Based on the thought above, add priority traversal conditions in the process of searching in the bounding box. Comparable of distance of centre before the traversal, traverse according to distance from the principle of the priority.

In the process of traversing bounding box tree, calculate the distance between the bounding boxes, first traverse bounding boxes in accordance with the short distance. If the short distance of the bounding box does not intersect, it is not necessary for a long distance. If the short distances of the bounding box intersect, showing two objects collision, reducing the detection of collision time. If it is necessary to calculate the numbers of triangle bounding box, which are intersect, it will need to traverse the far distance bounding box, traverse the process according to the principle of distance priority.

As shown in Figure 3, when traversing, not to join the first traversal strategy, until traversing the node disjoint, the tree will proceed as follows in accordance with each branch. Join the distance priority traversal, the bounding boxes of short distance, which does not intersect, then stop traversal. That is to reduce the width of tree, as shown on the right. When the depth of the tree increases to a certain extent, it will improve the efficiency of the algorithm.

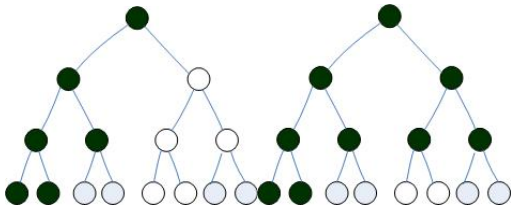


FIGURE 3. Traverse according to the distance priority strategy

2.5. STRATEGIES USING THE SPATIAL-TEMPORAL CORRELATION REDUCING TREE DEPTH TO ACCELERATE THE TRAVERSAL

From the above analysis, the key to improve the efficiency of the algorithm is to reduce the number of bounding box intersection test. The objects in the virtual environment motion continuously, dense temporal sampling point makes the change of object location and status in a period of time very small. This feature is called Spatial-temporal correlation. Time correlation object leads to spatial correlation. If object collision does not occur in the last time, it is not likely collide in the current time. If two objects collide at a last point time, it is also likely to collide in the current point time. According to this characteristic, in order to improve the efficiency of the algorithm, use the spatial-temporal correlation traversal tracking strategy, and determine the traversal path at the present time by tracking tree traversal process the active object point in the environment object.

2.6. STRATEGIES USING THE SPATIAL-TEMPORAL CORRELATION REDUCING TREE DEPTH TO ACCELERATE THE TRAVERSAL

From the above analysis, the key to improve the efficiency of the algorithm is to reduce the number of bounding box intersection test. The objects in the virtual environment motion continuously, dense temporal sampling point makes the change of object location and status in a period of time very small. This feature is called Spatial-temporal correlation. Time correlation object leads to spatial correlation. If object collision does not occur in the last time, it is not likely collide in the current time. If two objects collide at a last point time, it is also likely to collide in the current point time. According to this characteristic, in order to improve the efficiency of the algorithm, use the spatial-temporal correlation traversal tracking strategy, and determine the traversal path at the present time by tracking tree traversal process the active object point in the environment object.

As shown in Figure 4, the root node is set to the initial node. Traverse the bounding box tree node, which is disjoint and record this node in the first time. Then traverse again from the node disjoint. To reduce the depth of bounding box tree, and decreases the number of bounding box intersection test, improve the efficiency of the algorithm. The correctness of the algorithm is secure, because from the root node to the tree node is surrounded by only a tag node, and all nodes labelled as complete, the intersection is empty.

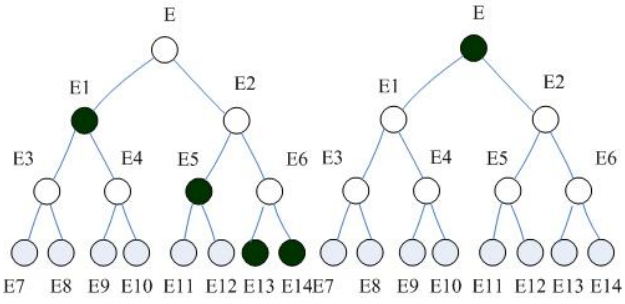


FIGURE 4. Schematic diagram of traversing bounding box tree by the tracking tables (green nodes were labelled nodes)

There are three types of labelled nodes, internal disjoint node, disjoint leaf nodes and leaf nodes of intersection. In the worst case, mark the node to all the leaves when the environment object and object are likely to occur coincide.

Traversal using temporal correlation, algorithm is as follows:

Let vT be the first marker node of tracking table, vF as the active object root node.

1) If the vT is not empty, then judge the bounding box vF and vT and judge whether intersect. If it intersects into 2), else enter the 5).
2) If the bounding box intersects, judging whether the vT is a leaf node. If it is a leaf node, then enter the 3) judgment. Or into the 4).
3) If it is a leaf node, then traverse the corresponding bounding box tree.
4) If not a leaf node, the node is deleted, down and update. For each sub of vT nodes to determine whether an internal node with vF which is disjoint intersect. At the current time if intersecting, insert this node from tracking table, and mark the new node.
5) Achieve vN as the next node from tracking table, determine vN and vT as the brother nodes. If it is the brother node, judge whether the bounding box of vN and vF intersect.
6) If the bounding box is disjoint from the track, and delete the vT and vN, insert the parent node into the current position.
7) vT points to the next node traversal, until complete tracking table.

The principle of updating the trace table is that, if an internal node in the trace table joins in the last moment and disjoint in the present moment with the root node, delete this node from the tracking tables, and add the node tree as a marker of node traversal. If the table has two brothers for nodes in the current situation and the root node disjoint, then delete the brother node from the table, with the parent nodes for replaced.

Below is the example of using of a tracking list traversal, using depth first traversal search. Two fork tree representation environment object bounding box tree. There are 8 leaf nodes in the graph (E7~E14), assuming that each leaf node contains only one basic geometric elements, E is the root node. The FIGUREURE below illustrates the case of collision sampling point when active object after environmental object at four consecutive time. Solid nodes represent moving object tracking in the table's contents at all-time sampling point, called marked node. It describes the traversal process where the root node of active object in environment object bounding box tree.

At t_i time (see Figure 5), arriving at the internal nodes of E1, E5, E6 find that detect bounding boxes are disjoint. Therefore, there is no need to continue the search. The node disjoint recorded as mark point.

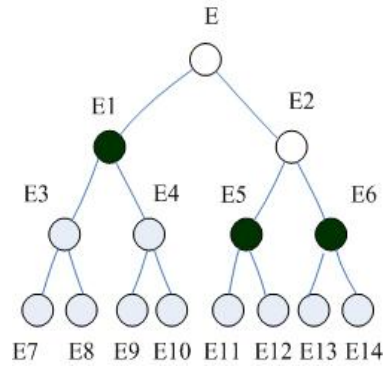


FIGURE 5. Searching result when time= t_i, trace table= (*E1*,*E5*,*E6*)

At time t_{i+1} (see Figure 6), if you do not use the tracking list traversal strategy, traversal path is $E \leftarrow E1 \rightarrow E2$. There is 3 times bounding box intersection test. After the use of tracking strategy, traversal path is $E1 \rightarrow E5 \rightarrow E6$. They are also need for three intersecting detection.

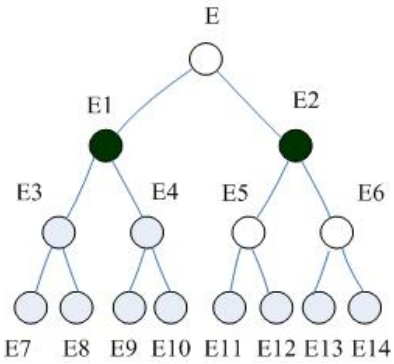


FIGURE 6. Searching result when time=t_{i+1}, trace table =(E1, E2)

At time t_{i+2} (see Figure 7), if you do not use tracking strategy, traversal path is $E \rightarrow E1 \rightarrow E3 \rightarrow E4 \rightarrow E2 \rightarrow E5 \rightarrow E6$, there are 7 bounding box intersection tests. The tracking strategy traversal path is $E \rightarrow E3 \rightarrow E4 \rightarrow E2 \rightarrow E5 \rightarrow E6$.

They need 6 bounding box intersection tests.

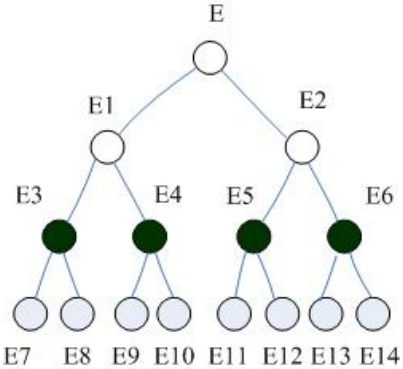


FIGURE 7. Searching result when time=t_{i+2}, trace table=(E3, E4, E5, E6)

At t_{i+3} time (see Figure 8), if you do not use the tracking strategy, path is $E \to E1 \to E3 \to E7 \to E8 \to E4 \to E2 \to E5 \to E11 \to E12 \to E6.$

The path of using the tracking strategy is $E3 \to E7 \to E8 \to E4 \to E5 \to E11 \to E12 \to E6$.

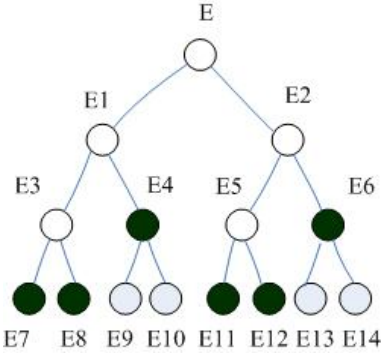


FIGURE 8. Searching result when time = t_{i+3}, Trace table = (E7, E8, E4, E11, E12, E6)

There are only need 8 bounding box intersection test. Table 1 gives the comparison of bounding box intersection number by tracking strategy.

TABLE 1. Comparison of bounding box intersection number

STRATEGY\TIME	t_{i+1} time	t_{i+2} time	t_{i+3} time
do not use the tracking strategy	3	7	11
use the tracking strategy	3	6	8

From the table we can see, along with the increase of the depth of the tree, the tracking effect is more obvious. When the depth of the tree increases to a certain degree, the tracking strategy will accelerate algorithm.

2.6.1. The improvement based on spatial-temporal correlation

1) When the intersection degree between two objects is small, the spatial-temporal correlation of traversal does not have the advantage. Although reducing the depth of the tree, there are need to add additional memory to store trace table space. The effect is not obvious. Therefore, this paper proposes the initial bounding box distance to determine whether using a trace table optimization strategy. When the distance of two bounding box is less than a certain value, use the tracking table optimization strategy.
2) When the objects completely intersect, tracking table records all the leaves, which will increase the length of the trace table. When the trace table length reaches at 2564 by the test, the memory allocation problems will cause crashes. (Limited to the experimental environment, the data for reference only) so you can set the maximum length of the trace table. When the trace table is greater than a certain value, use the traversal strategy directly, no longer down update tracking table.
3) As the distance of two bounding box increases, the bounding box not to update. Each need a lot of elements from the following table, but there are few nodes from the root node, which can determine the bounding box does not intersect. However, the use of traversal strategy would increase the storage space of a memory, reducing the efficiency of traversal. In order to avoid the occurrence of the above, it is used when the distance between the bounding boxes is reduced.
4) For the algorithm proposed in 2.4, generate a new tree, they are need for the whole tree according to the distance, add storage space. It also has complete traversal after the operation in the process of generating new tree. If only sort for the tracking elements in the table by distances, is equivalent to add index for the nodes on the tree. We sort by the index through distance. When returning disjoint distance index, there is no need to traverse the distance index. In solving the index table sorting problems, to abandon the use of linked list storage, use of two binary heaps.

Because when using a linked list for the N tag node sort, time complexity is O (N^2), read time is O (1), the total time complexity is O (N^2) +O (1). For the N tag node construction of pile, the time complexity is O (n), the average time per operation complexity and the worst case time complexity is O (log n), the total time is O (n log n). For the two fork heap, the total time complexity is O (n log n) +O (n) =O (n log n), fast than the list.

3. Test Results and Analysis

We chose the model of Stanford University as a data validation. Experiments using VC++ run on a PC, which CPU is Pentium (R) Dual-core, memory is 2G. To visualize the reality of collision scenarios, use OpenGL display the data. We use two experimental scenarios, increasing the number of triangular slices. Scene **1** (see Figure 9) is rabbit model, containing 16,301 triangles slices.

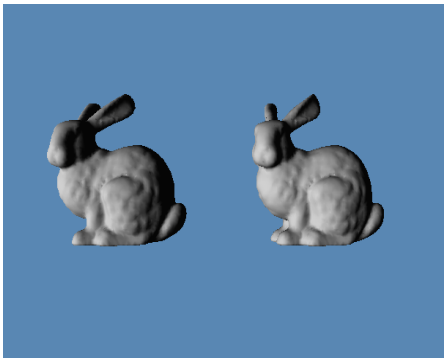


FIGURE 9. Scene 1 (rabbit model)

Scene 2 (see FIGURE 10) of the dragon model contains 43,765 triangular slices.

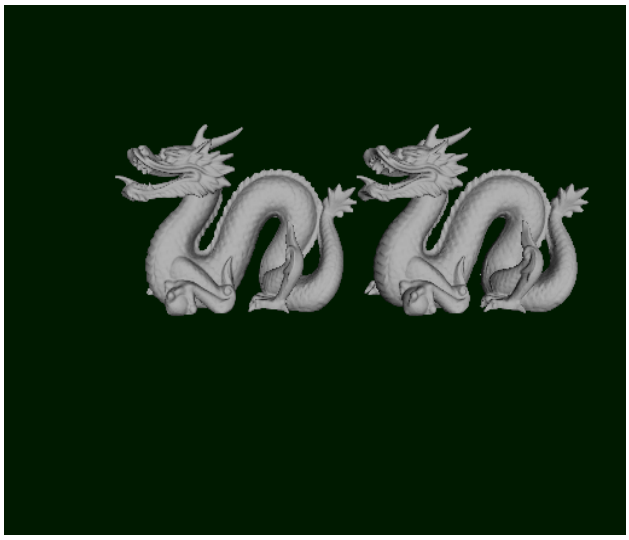


FIGURE 10. Scene 2 (dragon model)

Compare the number of bounding box using prediction algorithms as shown in TABLE 2. Based on the above data, drawing graphics are shown in Figure11.

TABLE 2. Comparison of bounding box intersection number:

A	59	467	9001	13677	15203	18907	33063	36585
B	49	365	7729	11661	12849	15663	24149	30959

A *- number of bounding boxes intersect, which don't use Prediction algorithm,*
B*- number of bounding boxes intersect which use Prediction algorithm*

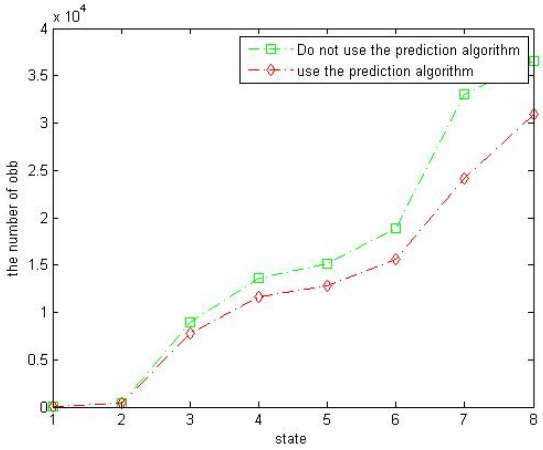


FIGURE 11. Compare the number of the bounding box diagram, which contains Prediction method for testing and the method of separating axis.

The abscissa represents different states of the bounding box; the ordinate axis represents the intersection number of the bounding box. Blue line represents the number of bounding box, which does not use prediction method. The green line shows the number using the prediction method. 1-8 states represent two objects from far and near in the process. In this experiment, State 1 represents the x axis distance of 1; the other state represents translation reducing distance 0.1 than the previous status. Comparison of the bounding box using tracking Policy is shown Table 3.

TABLE 3. Comparison of bounding box intersection number:

A	5	9	59	467	9001	13677	15203
B	5	7	55	438	8768	8507	8009

A *- the number of bounding boxes intersects which don't use Tracking strategy,*
B *- the number of bounding boxes intersects which use Tracking strategy*

For better performance of the law, according to the above data, drawing graphics are shown in Figure12.

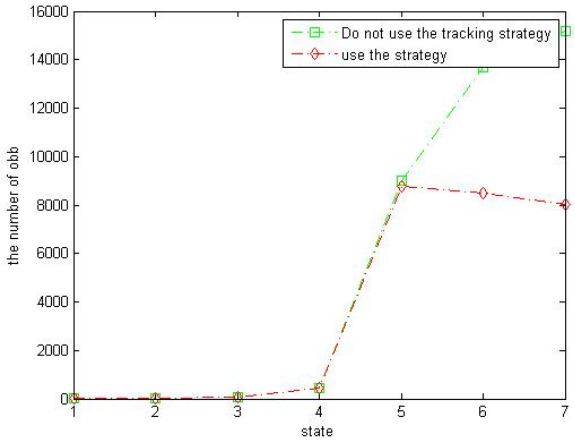


FIGURE 12. The number of the bounding box using Tracking Policy

The abscissa represents different states of the bounding box; the ordinate axis represents the intersection number of the bounding box. Green line represents the number of bounding box, which does not use the tracking policy. The red line shows the number using the policy. For better performance of the law, seven states are selected. State 1 indicates that two objects abscissa distance of 1.2.The other state represents translation reducing distance 0.1 than the previous status. From the above chart, we can see that the use of tracking optimization strategy more obvious effects in the case that the bounding box tree reaches a certain depth.

As is shown in FIGURE 13, the time period uses a clock count value by dividing the clock cycle, only as a measure. Similarly select eight states. Which 1-8 states represent two objects from far to near .State 1 represents the x axis distance of 1, and the other state represents translation reducing distance 0.1 than the previous status.

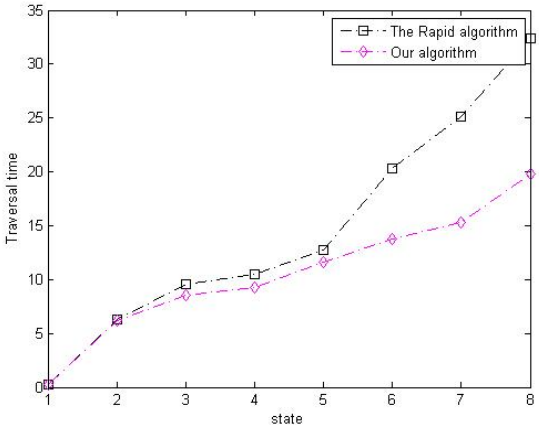


FIGURE 13. Comparison of average traverse time

4. Conclusions

From discussed above, we reduce the test number of detection of bounding boxes by forecasting method to improve the efficiency of the approach and the temporal and spatial correlation which reduces the traversal depth and width of tree, so as to improve the efficiency of the algorithm. Verified by experiments, this algorithm hold more efficient than which of OBB in collision detection. According to the experimental results above, this paper, mainly focus on the collision of two objects. In future research, we will study how to improve the collision detection of multi objects, to accelerate the traversal algorithm through utilize the characteristic of spatial-temporal correlation. We can also use a trace table tracking, which has a limited length to increase the efficiency of this algorithm. The parallel algorithm can be used for attaining this proposes as well.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.61370141, 61300015), by Program for Changjiang Scholars and Innovative Research Team in University (No.IRT1109), the Key Project of Chinese Ministry of Education (No.209029), the Program for Liaoning Excellent Talents in University (No.LR201003, LJQ2013131), the Program for Liaoning Science and Technology Research in University (No.LS2010008, LS2010179), the Program for Liaoning Innovative Research Team in University (Nos. LT2010005, LT2011018), Natural Science Foundation of Liaoning Province (201102008, 2013020007) and by "Liaoning BaiQianWan Talents Program (2010921010, 2011921009).

References

[1] Goldsmith J., Salmon J. (1987) *Computer Graphics and Applications IEEE,* **7**(5), 14-20
[2] Palmer I. J., Grimsdale R. L. (1995) *Computer Graphics Forum Blackwell Science Ltd,* **14**(2), 105-116
[3] Kim Y. J., Lin M. C., Manocha D. (2004) *Algorithmic Foundations of Robotics V*. Berlin Heidelberg: Springer, 505-22
[4] Carmona R., Navarro H. (2012) *Int. Journal of Creative Interfaces and Computer Graphics*, **3**(1), 16-27
[5] Lauterbach C., Mo Q., Manocha D. *Computer Graphics Forum,* **29**(2), 419-428
[6] Pabst S., Koch A., Straßer W. (2010) *Computer Graphics Forum,* **29**(5), 1605-12
[7] Huang R. (2012) *Proc. of Proceedings of the IEEE Symposium on Electrical & Electronics Engineering (EEESYM'12),* Kuala Lumpur, 589-91
[8] Tu C., Yu L. (2009) *First International Workshop on Education Technology and Computer Science (ETCS'09),* Wuhan, Hubei, 331-3
[9] Chang J.-W., Wang W., Kim M.-S. (2010) *Computer-Aided Design,* **42**(1), 50-7
[10] Xie Q.-R. and Geng G.-H. (2012) *Journal of Computer Applications,* **32**(3), 719-721
[11] Chang J-W, Kim M-S (2009) *Computers & Graphics,* **33**(3), 235-40
[12] Fan L. M. (2012) *Advances in Computer Science and Information Engineering* Heidelberg: Springer, 589-94
[13] Hu T-dong (2011) Information Computing and Applications: Springer, 602-9
[14] Ying-Mei W. (2000) Changsha: *National University of Defense Technology,* (in Chinese)
[15] Guo L., Zheng Y., Liu J. (2013) *Computer Applications and Software,* **30**(5), 174-6
[16] Erbes R., Mantel A., Schömer E. (2013) *Facing the Multicore-Challenge II*. Heidelberg-Berlin: Springer, 84-95
[17] Zhang X., Kim Y. J. (2007) *IEEE Transactions on Visualization and Computer Graphics,* **13**(2), 318-29
[18] Redon S., Kheddar A., Coquillart S. (2002) *Computer Graphics Forum, 23rd Annual Conference of the Eurographics Association Saarbrucken Germany* Saarbrucken: Blackwell Publishing Inc., 279–87
[19] Redon S., Kim Y. J., Lin M. C., Manocha D. *Proceedings of ACM Symposium on Solid Modeling and Applications (Eurographics Association).* Genova*:* ACM Press, 145–56

Authors

Ranran Man, born 22.02.1991, Zaozhuang Shandong, China

University studies: She is currently learning for master's degree in Key Laboratory of Advanced Design and Intelligent Computing, Dalian University
Scientific interest: The main research direction is computer animation, and the research focus is collision detection

Dongsheng Zhou, 23.10.1978, Yuncheng Shanxi, China

Current position, grades: vice professor in Key Laboratory of Advanced Design and Intelligent Computing, Ministry of Education at Dalian University
University studies: Ph.D. degree in mechanical design from Dalian University of Technology in 2010
Scientific interest: His research interests include computer graphics, computer animation, and data mining and intelligent algorithm. He is a member of ACM and CCF

Qiang Zhang, 30.08.1971, Xi'an, Shaanxi, China

Current position, grades: Now he is a professor at Dalian University, Dalian, China
University studies: Ph.D degree in Circuits and System from Xidian University in 2002
Scientific interest: His research interests are computer animation, intelligent computing.
Publications: Now he has served as editorial boards of seven international journals and has edited special issues in journals such as Neurocomputing and International Journal of Computer Applications in Technology

Precision agriculture compressed sensing and data fusion algorithm for wireless sensor networks

Liu Hong-Xia[1*], Zhang Feng[1, 2]

[1] *School of Information Engineering, Yulin University, 719000, Yulin, China*

[2] *School of Software, Xi'an University, Xi'an 710065, China*

Received 1 January 2014, www.tsi.lv

Abstract

In order to improve the energy efficiency of WSN nodes and prolong the life of the network, reduce data redundancy. In this paper, proposed the spatial correlation node data compression and fusion algorithm based on the theory of compressed sensing. Firstly, make signal node random projection based on time correlation, then, for random routing instability and network transmission of data fusion technology reconstruction energy effects, proposed the energy consumption of compressed sensing and clustering data fusion technology. The experiment showed that after data fusion, not only effectively removes the redundant information of neighbouring nodes, and the reconstruction error is small, and can accurately realize data decompression, thereby reducing the node communication in wireless sensor network capacity, reduce power consumption of node, provides important support for the field environment in large scale wireless sensor network deployment.

Keywords: compressed sensing, intelligent wireless sensor, agriculture IoT, automatic control, data fusion.

1. Introduction

In recent years, the Internet of Things (IoT) as an international research hotspot, have obtained broad attention [1]. It's represents the future trend of development of the network, and requires sharing interoperability and information, so as to realize human society, the information space, the physical world ternary comprehensive connectivity and integration as the goal. Therefore, the Internet of things is regarded as the third technological revolution in information field [2].

In WSN, due to the large number of sensors, individual sensor nodes with limited resources and processing capacity and have highly redundant data, transmit the data to sink node separate method is not reasonable in the collection of information in the process of using each node [3]. To consider the accurate degree of data, network performance, resource consumption and reliability constraints, solve the associated data between nodes, sharing and integration. The fusion performance in wireless sensor networks, the algorithm should be simple, easy to implement, and reduce resource and energy costs.

An important feature of WSN network is data-centric, data processing is a key problem in WSN, largely determine the performance of the network. WSN nodes are generally battery-powered node energy is limited. How to save node energy and prolong the life of Yes WSN research is a key issue. In this paper, a large-scale and high density WSN as research background, for its data collection, transmission, storage process, the problems, the introduction of compressed sensing theory, the focus of research networks in the process of collecting large amounts of data present in large amount of data problems. In precision agriculture WSN, the general area of relatively is large, relatively large amount of information acquisition.

2. Compressed Sensing Based Spatial Correlation of Distributed

2.1. COMPRESSED SENSING THEORY

Compressed sensing or compressed sensing (Compressive Sensing, CS) is a theory developed in recent years, is a new data acquisition theory. Innovation is the core of the theory: for compressible data signals can be much lower than the Nyquist sampling via a standard way, is still able to accurately recover the original signal. This theory is largely broadened the scope of the compressing signal, in the theoretical framework for data integration CS, since the data amount is increased, the amount of data transmission can be effectively reduced, thereby reducing the communication energy [4].

Compressed sensing theory mainly includes three aspects: sparse signal representation, measurement selection matrix remodelling achieve optimal algorithm. The theoretical framework can be represented in Figure 1.

* Corresponding author. Tel: +86-189-9220-3580;
E-mail: 75419305@qq.com

First, if the signal $x \in U^N$ is compressible in an orthogonal basis ψ , get $\Theta = \Psi^T X$ is determined Ψ transform coefficients or the equivalent approximation sparse representation;

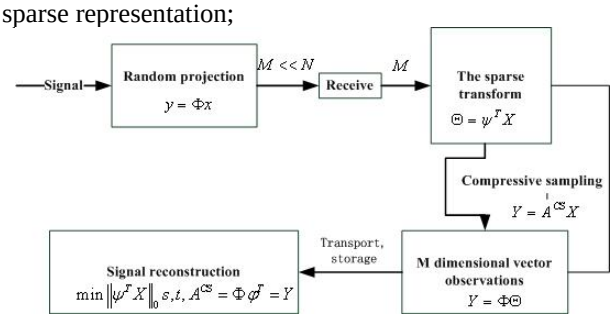


FIGURE 1. Theory framework of CS

the second step, a smooth design, and not related to Ψ transform base $M \times N$ dimensional observation matrix ϕ of observations Θ obtained observations $Y = \Phi\Theta = \Phi\varphi^T$ collection, the process may also be expressed as a non-adaptive observation $Y = A^{CS} X$ signal X (where $A^{CS} = \Phi\Psi^T$) @ through the matrix A^{CS} , the information is called CS operator; Finally, the 0 – norm optimization problem solving exact value of X or myopic approach $\hat{X}$:

$$\min \left\| \psi^T X \right\|_0 s,t, A^{CS} = \Phi\varphi^T = Y . \quad (1)$$

At present the theory of compressed sensing research is still in the initial stages, but already in some areas the signal and image processing, bio-sensing, analogy information conversion showed great potential. Currently compressed sensing theory research focuses on theoretical and mathematical reasoning to improve the level of the two most be studied. In a study of wireless sensor networks, compressed sensing theory has been the concern of more and more scholars, and related researchers. GHua and U.Chen take advantage of a new source coding techniques to improve the effectiveness of the relevant data. J.Haupt and his colleagues on the use of compressed sensing data acquisition and compression carried out relevant research [5]. Distributed compressed sensing theory concepts through joint sparse data on the correlation between the internal and the data are explored.

2.2. DISTRIBUTED COMPRESSED SENSING

The theory of CS standard is mainly for the design of the internal data structure of the single node, namely the correlation data aware of individual nodes to reduce the node data dimension. In wireless sensor networks, a large number of sensor nodes. For the node space intensive distribution, there is existence of high correlation between adjacent nodes of spatial data. Using this characteristic, natural to think of each node collaboration form measurements are combined, to reduce the overall dimension of data measured by correlation between node data compression effect, will get more. Based on this, the theory of distributed compressed sensing (DCS) emerges as the times require. D. Baron et al. Based on the theory of CS further extends the application of CS, put forward a distributed compressed sensing (DCS, Distributed Compressed Sensing) concept gas to compress a single signal sampling was extended to compress the signal group sampling [7]. The establishment of DCS theory is based on a "joint sparse signals known as group model (JSM, Jointly Sparse Models)" concept [8]. DCS theory pointed out that, if a plurality of signal are8 able to express the sparse in the same medium, and these signals is also has the correlation, then every signal in coding can use another not associated with the sparse matrix based (such as a random matrix) for sensing data observation and coding, which is far less than the original length of the encoded signal.

2.3. SPATIAL CORRELATION IN DISTRIBUTED COMPRESSED SENSING MODEL BASED ON WSN

In the WSN to collect data through various Sink node sensing sensor nodes, and monitoring the event source S to estimate the area, making the distortion estimation results to meet the requirements of WSN applications. Figure 2 shows the compressed sensing based spatial correlation of distributed.

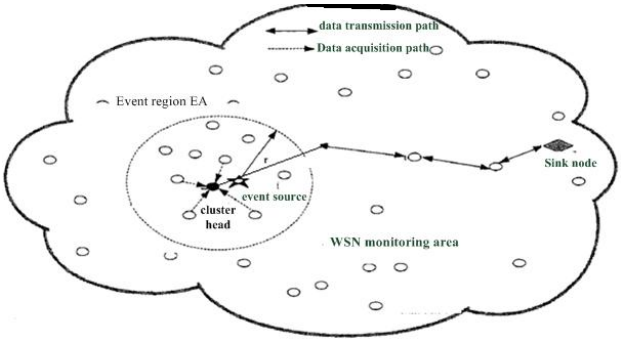


FIGURE 2. Schematic perceptions based on spatial correlation of distributed compression

As can be seen from Figure 2, because the WSN nodes are densely distributed in S_j and between S_j and S in the presence of varying degrees of spatial correlation. Therefore, to meet the requirements of distortion, by determining the scope of the EA event area, while taking advantage of the perception of spatial data between WSN node correlation data compression and reconstruction, has a very important significance in the WSN.

Assuming the event source S WSN location (0,0), the distribution area in the event of a node $EAn(x, y)$,

$n(x_r, y_r)$ coordinates are $n(x, y)$, $n(x_r, y_r)$, its perception data is $S(x, y)$, $S(x_r, y_r)$ variorums is defined as:

$$\gamma(x) = \frac{1}{2}\left[(S(x, y) - S(x_r, y_r))^2\right], \tag{2}$$

where $(x - x_r)^2 + (y - y_r)^2 = r_2$, if the value of the variorums $\gamma(x)$ is smaller, then the perceived correlation between the data will be stronger. In polar coordinates can be defined within a node $n(r, \theta)$ perception data to $S(r, \theta)$, EA within the scope of the event area in the monitored area WSN triggered the event source S , between the perception of sensory data node $n(0,0)$ data and information to other nodes around the following correlation:

$$\begin{matrix} S(0,0) = I_{(U=T)} + I_{(U=H)} \int_{\theta} \int_{\gamma} (S_{(r,\theta)} + Z) \delta(R = r) \\ \delta(\Theta = \theta | R = r) \end{matrix}, \tag{3}$$

wherein $U = T$ represents sensing data $S(0,0)$ is obtained by a random variable γ , the probability is β ; $U = T$ data through the sensing means adjacent to the $S(0,0)$ sensing data of the node $n(r, \theta)$ is worth to $S(r, \theta)$, the probability is $1 - \beta$. Random variable Y and Z are the probability density function $f_Y(y)$, $f_Z(z)$ and a random variable and random variables Y and Z are each independently sensing data $S(r, \theta)$.

3. Spatially Distributed Compressed Sensing Information Acquisition Method

3.1. GREEDY ALGORITHM

The basic steps of the greedy algorithm are:

1. Given initial estimates of $\theta^0 = 0$;
2. At each iteration, according to $A(\theta - \theta^0) = A\theta - A\theta^0$ determine the estimated value Δ of $\theta - \theta^0$;
3. Leaving only the larger Δ value items, other items will be set to zero. Update $\theta^0 = \theta^0 + \Delta$.

There are many ways about greedy algorithm, mainly has the orthogonal greedy algorithm (Orthogonal Matching Pursuit, OMP), regularized orthogonal greedy algorithm (Regularized Orthogonal Matching Pursuit, ROMP), piecewise orthogonal greedy algorithm (Stage wise Orthogonal Matching pursuit STOMP) and the gradient of the greedy algorithm (Gradient pursuit method).

3.2. ITERATIVE THRESHOLD ALGORITHM

Iterative threshold algorithm is a recursive method to determine a solution of $y = A\theta$, specifically, is defined by the following equation sequence θ^n to close to the optimal solution $y = A\theta$:

$$\theta^{(n+1)} = S\tau\left[\theta^{(n)} + A^* A\theta^{(n)}\right], \tag{4}$$

where $S\tau(x)$ is: $S\tau(x) = \begin{cases} x - \tau, & x > \tau \\ 0. & x \le \tau \\ x + \tau, & x < \tau \end{cases}$.

Gets the iterative sequence can be obtained by two methods: the operator splitting and two function approximation method. Firstly, by introducing a regularization factor K that set (P1) problem is transformed into unconstrained problems as follows:

$$\min L(\theta, \lambda) = \|\theta\|_1 + \lambda \|y - A\theta\|_2^2. \tag{5}$$

Because $\|\theta\|_1$ is non smooth convex separable function, $\|y - A\theta\|_2^2$ is a smooth non separable convex function, so if set $T_1 = \partial \|\theta\|_1$, $T_2 = \partial \|x - D\theta\|_2^2$, so we can write

$$\begin{aligned} &\hat{\theta} = \arg\min L(\theta, \lambda) \\ &\Leftrightarrow \hat{\theta} \in (T_1 + T_2)(\theta^*) \\ &\Leftrightarrow \hat{\theta} \in (I + \tau T_1 - I + \tau T_2)(\theta^*) \\ &\Leftrightarrow (I - \tau T_2)(\theta^*) \in (I + \tau T_1)(\theta^*) \\ &\theta^* \in (I - \tau T_2)^{-1}(I + \tau T_1)(\theta^*) \end{aligned} \tag{6}$$

The $(I - \tau T_2)$ is a step τ of the gradient operator, and $(I - \tau T_2)^{-1}$ is $S\tau(\bullet)$.

4. Spatial Correlation of Distributed Compressed Sensing Algorithm

First of all, need to get the distribution range of event region EA. At the position of the hypothetical event sources have a virtual node is $n(0,0)$, the information data of the node $S(0,0)$, $S(r, \theta)$ data boundary nodes in $n(r, \theta)$ aware S the events triggered by the event source region of the $|S(r, \theta) - S(0,0)| \le \mu$, μ is in line with the conditions, the error threshold, it is used to indicate the difference in perception data information and the event source at node in different position of the distribution of radius r represents the event region EA. According to equation (2) can be obtained:

$$r(r)=\frac{1}{2}E\left[\left(S(r,\theta)-S(0,0)\right)^2\right]=\frac{1}{2}E\left[Z^2\right]$$

$$=\frac{1}{2}\int_{-\varpi}^{\mu}z^2\frac{1}{\sqrt{2\pi\sigma_z}}e^{\frac{z^2}{2\sigma_z}}dz \quad , \qquad (6)$$

$$\frac{1}{2}\sigma_z^2 erf\left(\frac{\mu}{\sqrt{2\pi\sigma_z}}\right)-\frac{1}{\sqrt{2\pi\sigma_z}}e-\frac{z^2}{2\sigma_z^2}=\psi(\sigma_z,\mu)$$

where $\psi(\sigma_z,\mu)$ is a function of σ_z and μ, $erf(x)=\frac{2}{\sqrt{\pi}}e^{-t^2}dt$.

In summary, the distributed spatial correlation of compressive sensing encoding and decoding algorithm based on WSN can be expressed as follows:

Step 1: according to the formula of $T^J(\lambda_j)=\{\lambda_0,\gamma_0,\gamma_1,...,\gamma_{J-1}\}$, the distribution range of the sink node computing event region EA, forming a cluster by multicast routing way node $n_i(i=1,2,...,N)$ activation events in area EA.

Step 2: cluster head node n_h to generate the observation matrix $\Phi=R(S_M,\tau_N)$, where $R(\bullet)$ is a pseudo random number generating function $\tau_N:\{\tau_1,\tau_2,...,\tau_N\}$, τ_i is the number of n_i.

Step 3: the cluster head node of the coded data Y_M transmission by wireless multi hop to the sink node.

Step 4: Sink node and cluster head node generates the same observation matrix $\Phi=R(S_M,\tau_N)$, then get a small wavelet transform matrix ψ according to the network topology, run the distributed compressed sensing decoding algorithm.

5. Simulation and Performance Analysis

5.1. EXAMPLES OF DATA FUSION

Introduced in front of cluster Limited ad hoc networking mode, it is in accordance with the region of the cluster, the cluster head using ad hoc network between the nodes in the cluster structure, with star, tree and chain structure.

Assuming a field of soil moisture content within the region more than 45% of the area number for inquiries first sends a query request. Suppose there are four field using tree routing, the sensor position and the communication path shown in Figure 3.

In Figure 3, each sensor node has prepared a data, and (Filed, Soil moisture) is expressed in the form of. Need to do the following:

1) Check each sensor node is in accordance with the upload request, to decide whether to participate in the transmission;

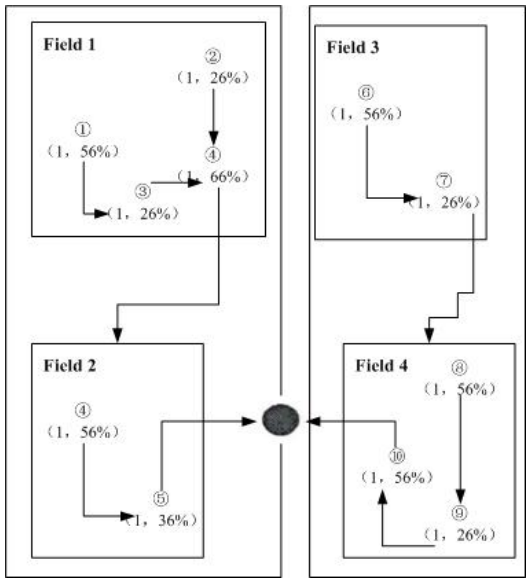


FIGURE 3. Instance of data fusion

2) Each node receives the data sent by other nodes, for local operations, results of operations continue to submit to the upstream node;

3) intermediate nodes if in a certain period of time has not received the neighbour node to send data, then the default from neighbour nodes below, did not submit the data nodes request.

According to the above operation, the appeal processes a total of 5 copies in the network transmission of data. As shown in Figure 3, the Node10 sent 3 copies of the data, Node 8 to send 3 copies of data, Node 3 to send 1 copy of data. If you do not have any data fusion method, each individual node send data to the sink node, and the sink node concentration calculation results, the network data transmission quantity is 21, far greater than 6, the corresponding node energy consumption will significantly increase.

5.2. EXPERIMENTAL AND SIMULATION RESULT

Simulation research and focus on the relationship between the number of observation data and reconstruction error of DCS encoding and decoding algorithm of the time correlation analysis based on WSN. The simulation uses Matlab as tools, the spatial correlation of data source in WSN by using two-dimensional Gauss distribution to model, the node in WSN random uniform distribution of events in the 60mx60rn area. In the system, all the nodes of the sensor that produced in WSN sensing data at the same time and send it to the cluster head node encoding, the cluster head nodes using DCS coding algorithm on the perception of the received data compression, and the observation data generated by multi hop wireless transmission to the decoder, accurate reconstruction of the sink node using the DCS decoding algorithm on each node sensing data in the area of the WSN event.

As shown in Figure 4, the relationship between the number of simulation graph reconstruction error and observation data, assuming that perceived value number is 3000. Can be seen from the chart, the observed values of M increases, the reconstruction error is more and more and more small, suitable for multiple observations data acquisition system.

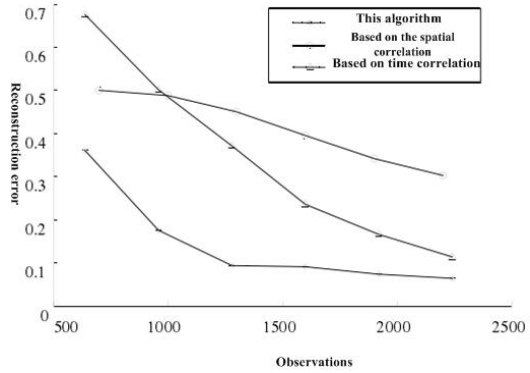


FIGURE 4. the relationship between the observed number and reconstruction error

6. Conclusions

Compressed sensing is a theory developed in recent years, firstly, the theory of compressed sensing do a simple introduction, focusing on the spatial correlation in WSN based on distributed compressed sensing theory. WSN in precision agriculture, the amount of information needed to capture the scene is relatively large, and therefore the number of sensor nodes required range. Is also a great amount of data transmission, data compression is an effective way to reduce the amount of data transmission, according to the practical application of this article, the choice of the spatial correlation based on compressed sensing theory of distributed network layer data compression.

Acknowledgments

This work is partially supported by the Funding Project for Department of Education of Shaanxi Province in 2012 #12JK0828, 2013 #2013JK1151 and 2013JK1167. Research and Cooperation Project for Department of Yulin city of (**#Sf13-08，#Sf13-23, #Ny12-19, #Ny13-10,#2012cxy3-26**). Thanks for the help.

References

[1] Pierce F. J., Elliott T. V. (2008) *Computers and Electronics in Agriculture*, **61**(1), 32-43
[2] Wetter M. (2011) *Journal of Building Performance Simulation*, **4**(3), 185-203
[3] Rodriguez R. I., Jia Y. (2011) *International Journal on Smart Sensing and Intelligent Systems*, **4**(2), 325-37
[4] Haifeng D., Xiaolong Z. (2012) *International Journal on Smart Sensing and Intelligent Systems*, **5**(4), 824-42
[5] Geziei S. (2008) A survey on wireless position estimation. *Wireless personal Communications*, **44**(3), 263-80
[6] Zhibin S., Chao T., Jun C., et al. (2010) *IEEE Transaction on Communications*, **58**(2), 511-20
[7] Candès E. (2006) Compressive sampling. *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, **3,** 1433-52
[8] Candès E. (2008) The restricted isometry property and its implications for compressed sensing. *C.R.Math.Acad.Sci.* Paris **346**(9-10), 589-92.

Authors

Liu Hong-Xia, born 24.10.1981, in Shannxi Yulin

Current position, grades: lecturer in Yulin University
University studies: MS degree in specialty of educational technology from Shaanxi Normal University in 2007
Scientific interest: Information technology education, application of Internet of things

Zhang Feng, born 26.06.1980, in Shannxi Yulin

Current position, grades: associate professor in Yulin University
University studies: MS degree in Computer science from Xidian University in 2009
Scientific interest: Cloud integrated manufacturing technology, the modeling of complex systems, the Internet of Things applications

Structural design and active control simulation for double beam actuator based on ANSYS

Yun-Feng Wang [1*], Wei Cheng [2]

[1,2]*School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, XueYuan Road No.37, HaiDian District, BeiJing, China*

Received 1 January 2014, www.tsi.lv

Abstract

A double-beam actuator (DBA) is designed in this study for attenuating the harmful vibration. Its finite element model including the folded beam, glue layers and piezoelectric laminar is developed in commercial software ANSYS. DBA's dynamic characters are analyzed by employing modal and transient analysis in ANSYS. The calculated results are used to develop an explicit state space model by using the observer/Kalman filter identification (OKID) technique and Eigen-system Realization Algorithm (ERA). The robust H_2 controller is designed based on the identified state space model and it is then incorporated into the ANSYS finite element model to perform the close loop controlling simulation. The results of simulation show the settling time reduce to 0.06s with active control under impulse exciting and the response amplitude decrease 20 dB under sine exciting, which demonstrate the validity of DBA in application of active control.

Keywords: active control, system identification, active control, ANSYS, piezoelectric

1. Introduction

Active vibration control is an important problem in structures. One of the ways to tackle this problem is to make the structure smart, intelligent, adaptive, and self-controlling by making use of one type of materials called as smart material as sensors and actuators [1]. Piezoelectric ceramic as one kind of these smart materials has many advantages such as low price, high reliability, and fast response. So it has been used in many vibration control researches [2]. Cantilever beam being the controlled plant is used in most of these studies with piezoelectric patches being glued on its surface as the actuators and the vibration level are measured by anther piezoelectric patch or strain gauge. Some researchers are reviewed as follows.

Manning made the flexible beam's vibration being controlled by employing the system identification technique and pole placement control method in [3]. The beam and piezoelectric patches' finite element model of the structure is developed and the closed loop control is employed in [4]. Singh [5] used the beam and piezoelectric patches' finite element model too, but the control law applied is modal strategies.

Xu [6] established the intelligent beam's finite element model and then the state space model was developed based on modal analysis in ANSYS. Control system design was developed by employing the toolbox of MATLAB. The influence of sensor/actuator location was studied and the results indicate that the location near to the clamped end is better for vibration control.

In [7] the efficiency of the control loop in software ANSYS has been validated by comparing the simulation results with the analytical for two degrees of freedom system, then the scheme, developing model and controlling simulation is applied to the active vibration control problem of the smart structures. It is observed that this scheme can successfully be applied for the smart structures.

Xing-Jian Dong [8] used the OKID approach to develop an equivalent linear model from the outputs of the commercial finite element software ANSYS. LQG control technique has been used to design control laws. Numerical results are presented to demonstrate the efficiency of the proposed scheme in simulating an actively controlled piezoelectric structure. Finally, a complete active vibration control system has been set up to conduct experimental investigation. From the experimental results, it is observed that satisfactory performance of vibration attenuation can be achieved.

In this study, a double-beam actuator (DBA) for vibration control is designed and analysed. In section 2, the structure and functions of DBA are presented. In section 3, the finite element model of DBA is developed and its dynamic characters are analysed by modal and transient analysis in ANSYS. In section 4, the explicit state space model is developed by employing the OKID technique and ERA algorithm in the context of MATLAB environment. In section 5, the robust H2 control method is used to design the active control law, then the controller designed is incorporated into the finite element (FE) model in ANSYS and the simulations for testing

* Corresponding author. E-mail: wangyunfengdoc@126.com

active control efficiency are evaluated under two diffident conditions. Concluding remarks on the performance of the DBA system are presented in Section 6.

2. Structure of DBA

Three parts are included in the DBA, the folded beam, piezoelectric patches, and glue layers. The four piezoelectric patches are used as actuators and sensors and they are glued on the two base beams as shown in Figure 1. Q Wang [9] has studied how the piezoelectric patch's location and shape affect its controllability and the conclusions show that the controllability index reaches its highest value when the size of the piezoelectric patch is designed to span the entire beam when the piezoelectric patches are collocated. Collocation of piezoelectric patches has some priorities, which are given in [10]. Thus, the same configurations are used in DBA. The upper and lower piezoelectric patches are actuators and the two middle patches are sensors. This DBA can work by single one with upper interface connecting disturbance loads and lower interface connecting plant, or some more DBAs can be assembled as a controlling platform for active control as shown in Figure 2 [11].

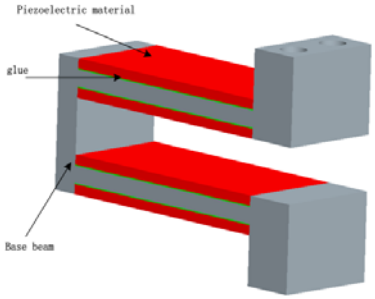


FIGURE 1. The structure of DBA Figure

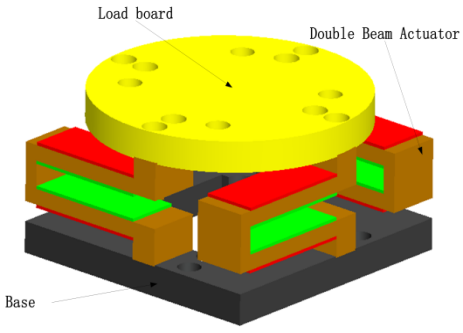


FIGURE 2. The structure of controlling platform

When external disturbance applies on the DBA, the sensors generate volt signals to measure the disturbance level and these signals are gathered by a computer. Controller designed analyses these signals and gives feedback control signals to actuators through a power amplifier. The process is shown in Figure3.

3. Finite Element Model Analysis

The finite element model of DBA is developed in ANSYS. The dimensions of the parts in DBA are shown in Table 1 and the parameters of materials in the finite element model are given in Table 2. In the FE model SOLID5 elements are used for piezoelectric patches while SOLID45 elements for folded beam and glue layers. SOLID5 element has an additional VOLT degree of freedom, which is the key for coupling the strain field and electrical field.

3.1. MODELING AND MODAL ANALYSIS

Modelling the DBA in ANSYS is as follows [12]. At first, the key points are located in the work plane, profile of DBA can be generated by connecting these key points. Then the profile is meshed by using SHELL43 elements, the finite element model of 3D elements is achieved by dragging the profile-area elements and then the materials' parameters and elements' properties must be assigned to the three parts of the model correctly. At the end, the constraints including degrees of freedom and electrical field are added to the model to complete the modelling process, all the degrees of freedom for lower interface area are set to be zero and the four piezoelectric patches' piezoelectric coupling relationships are set by CP command in ANSYS. The final model is shown in Figure 4.

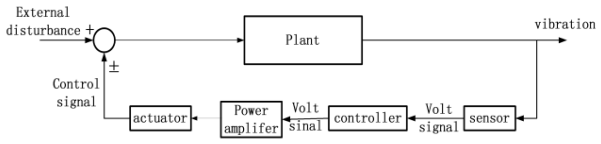


FIGURE 3. Flow chat of active control

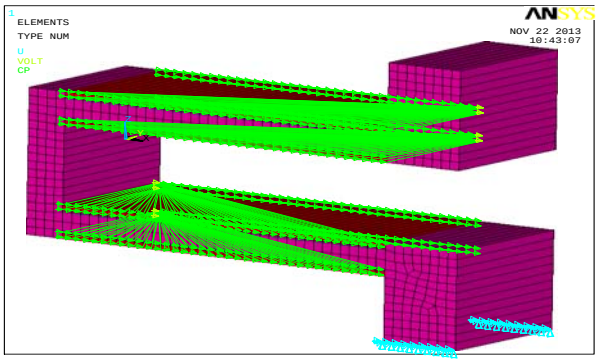


FIGURE 4. The FEM model of double beam actuator

Model analysis is performed to determine the first four steps natural frequencies. Only the reduced method

(Householder method) can be used for the FE model having coupled-field elements [12]. The results of modal analysis are given in Table 3 and the corresponding first two modal shapes are shown in Figure 5 and 6. It is found that the first modal shape is consistent with the upper beam's bending vibration and the second modal shape is consistent with the lower beam's bending vibration. The third and the fourth modal shapes are consistent to the torsion vibration of the upper beam and lower beam. Thus, the first two model shapes are the dominant vibration modes. Based on the results of modal analysis the Rayleigh coefficients of structural damping are defined as: $\alpha = 4e-6$ and $\beta = 2\alpha/3$. Time step is defined as $\Delta t = 1/(20 f_1)$ where f_1 is the first step natural frequency.

TABLE 1. The dimensions of DBA

Part	**Actuator**			**Piezo-electric laminar**	**Glue**
	High	***Length***	***Thickness***		
Dimension (mm)	42	60	20	45×20×0.8	45×20×0.2

TABLE 2. The parameters of materials in FEM model

Material	**Content**	**Symbol**	**Quantity**
Piezoelectric material (PZT-5)	Density(Kg/m^3)	ρ_p	767
	Dielectric constants(F/m)	Perx	8.93e-9
		Pery	8.93e-9
		Perz	6.92e-9
	Elastic compliance(m^2/N)	D11	1.3e-11
		D12	-9.1e-12
		D13	-9.1e-12
		D22	1.3e-11
		D23	9.1e-12
		D33	1.3e-11
		D44	4.35e-11
		D55	2.2e-11
		D66	9.1e-12
	Strain constants(C/N)	d31	-1.88e-10
		d33	6.7e-10
Aluminum	Density(Kg/m^3)	ρ_a	2700
	elastic modulus(Pa)	E_a	7.3e10
Glue	Density(Kg/m^3)	ρ_g	1100
	elastic modulus (Pa)	E_g	3.45e8

TABLE 3. The first four frequencies of the DBA

Step	**1st**	**2nd**	**3rd**	**4th**
Frequency (Hz)	416.09	666.84	690.91	1885.3

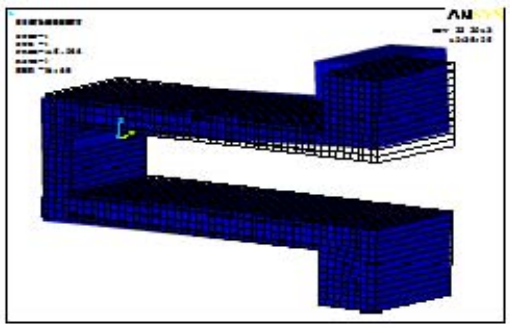


FIGURE 5. The 1st model shape of DBA

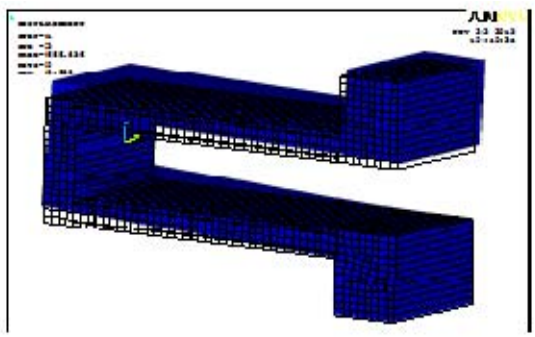


FIGURE 6. The 2nd modal shape of DBA

3.2. TRANSIENT ANALYSIS FOR MODEL

The control-system model of DBA is essential for closed-loop control simulation, but it is difficult to develop this model from the FE model directly in ANSYS. On the other hand identification technical which develops the control model based on the input-output data is much easier to implement. Before using the identification method enough input-output data must be existent, and these data can be generated from transient analysis in ANSYS.

Revoke the constraints on electrical field for all the piezoelectric patches before the transient analysis to simulate the open-loop condition. All the time steps are defined as $n_s = 1300$. A force $F_e = 10N$ is put on the upper interface of DBA at time $t = \Delta t$, and then this force is set to zero in the following time steps. The voltages generated from sensors are recorded and shown in Figures 7 and 8. It is seen that the output voltages' amplitude from the two sensors are different even though DBA is symmetrical in vertical direction. That is because the boundary conditions are different; the tail of upper piezo-beam is free whereas the lower piezo-beam's are clamped.

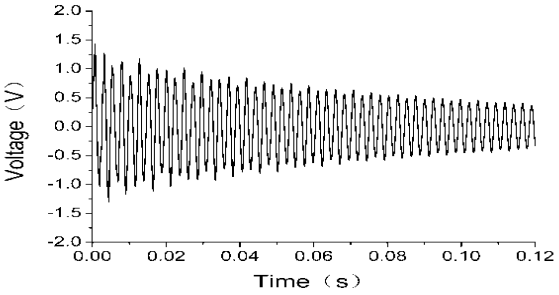


FIGURE 7. Output from sensor 1 under impulse disturbance

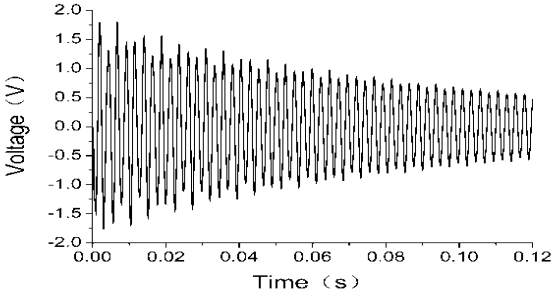


FIGURE 8. Output from sensor 2 under impulse disturbance

Remove the external disturbance and keep the actuator 2 and sensor 2 free. A voltage load $V_1 = 100V$ is applied on actuator 1 at time $t = \Delta t$. In the following time steps the voltage load is set to be zero and the voltage-outputs from sensor 1 are recorded and shown in Figure 9. The same operations are applied on actuator 2 while keeping actuator 1 and sensor 1 free and the voltage-output from sensor 2 are recorded too and shown in Figure10.

Because the two piezoelectric patches on the base beam are collocated and both of the collocated pairs can be treated as SISO systems, the coupling relationship between two piezo-beams is neglected [10]. The difference between the two output-voltages from sensor 1 and sensor 2 appear for the different boundary conditions as explained forenamed.

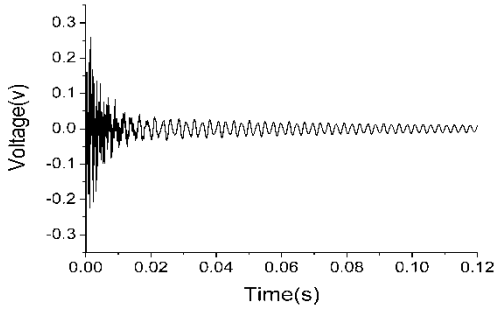


FIGURE 9. Output from sensor 1 under 100V impulse input to actuator 1

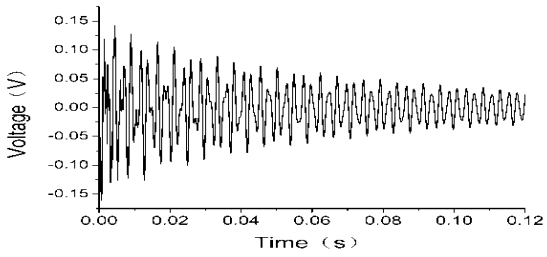


FIGURE 10. Output from sensor 2 under 100V impulse input to actuator 2

4. Control System Model

The parameters of linear time invariant system (LTI) can be identified from the OKID technical. One of the keys of OKID algorithm is the introduction of an observer into the identification process. The first step of identification process is the calculation of the observer's Markov parameters. Then the system's Markov parameters are determined recursively from the Markov parameters of the observer [13].

4.1. OKID IIDENTIFICATION TECHNICAL

The discrete state space form of a LTI system in time domain is given as

$$x(i+1) = Ax(i) + Bu(i), \tag{1}$$

$$y(i) = Cx(i) + Du(i), \tag{2}$$

where $A \in R^{n\times n}$, $B \in R^{n\times m}$, $C \in R^{q\times n}$, $D \in R^{q\times m}$. The relationship between input and output of this system can be rewritten as:

$$y(i)=\sum_{\tau=0}^{i-1}Y_\tau u(i-\tau-1)+Du(i), \tag{3}$$

where $Y_\tau = CA^\tau B$, Y_τ, and D are Markov parameters of system. If system (A,C) is observable, there is an observer as follows:

$$\hat{x}(i+1)=(A+MC)\hat{x}(i)+(B+MD)u(i)-My(i), \tag{4}$$

$$\hat{y}(i)=C\hat{x}(i)+Du(i). \tag{5}$$

When all the eigenvalues of $A+MC$ are zero, the estimated state $\hat{x}(i)$ converges to the real state $x(i)$ and (4) and Eq. (5) are rewritten as :

$$x(i+1)=(A+MC)x(i)+(B+MD)u(i)-My(i), \tag{6}$$

$$y(i)=Cx(i)+Du(i). \tag{7}$$

A system can be described by its input-output relationship as

$$y(i)=\sum_{\tau=0}^{n-1}\overline{Y}_\tau\left[u(i-\tau-1)y(i-\tau-1)\right]^T+Du(i), i\geq n, \tag{8}$$

where

$$\begin{aligned}\overline{Y}_\tau &= \left[C(A+MC)^\tau(B+MD)-C(A+MC)^\tau M\right]\\ &=\left[\overline{Y}_\tau^{(1)}\overline{Y}_\tau^{(2)}\right]\end{aligned}. \tag{9}$$

Then the eigen-system realization (ERA) algorithm can be employed to get the parameters of state-space form equation and the process is given as follows.
Construct the Hankel matrix $H(l,\tau)$, which includes $l\times l$ Markov parameters as

$$H(l,\tau)=\begin{bmatrix} Y_\tau & Y_{\tau+1} & \cdots & Y_{\tau+l-1}\\ Y_{\tau+1} & Y_{\tau+2} & \cdots & Y_{\tau+l}\\ \vdots & \vdots & & \vdots\\ Y_{\tau+l-1} & Y_{\tau+l} & \cdots & Y_{\tau+2l-2}\end{bmatrix}. \tag{10}$$

The order of the system is determined from the singular value decomposition of $H(l,0)$:

$$H(l,0)=U\sum V^T, \tag{11}$$

where U and V are normalized matrices. The parameters of discrete state space (1) and (2) can be achieved as:

$$A=\sum^{1/2}U^T H(l,1)V\sum^{-1/2}, \tag{12}$$

$$B=\sum^{1/2}V^T E_m, \tag{13}$$

$$C=E_q^T U\sum^{1/2}, \tag{14}$$

where $E_q^T=\left[I_q\ 0_{q\times(l-1)q}\right]$, $E_m^T=\left[I_m\ 0_{m\times(l-1)m}\right]$.

4.2. CONTROL MODEL OF DBA

The inputs to DBA model include the disturbance and control. The corresponding input matrices are defined as E and B. The complete system model of state space form for the DBA can be written as:

$$x(i+1)=Ax(i)+Bu(i)+Ed(i), \tag{15}$$

$$y(i)=Cx(i)+Du(i), \tag{16}$$

where A, $B=\left[B_1\ B_2\right]$, C and D preserve uniformity with (1) and (2). B_1 and B_2 are the input matrices of actuator 1 and actuator 2, E is the input matrix with dimension $n\times 1$ correspond to disturbance. Based on the transient analysis before, impulse disturbance and response data from the two sensors are used to identify the matrices E by employing OKID and ERA methods. In the same manner, matrices A, B and C can be identified by making use of the excitation and response data while voltage impulse input applying on the actuator 1 and actuator 2. By employing (3) D can be got as $D=y(0)/u(0)$.

In order to check the validity of the identified model, the frequency response functions (FRF) of the identified model and ANSYS model are obtained and they are compared as shown in Figures 11-14. Agreement is fairly good between the four results except for the frequency range above 2500Hz. In most conditions only the first few modes play dominant parts in the vibration response of the system, thus it can be concluded that the state space model developed by identification technique characterizes the dynamics of system and can be used for control law design.

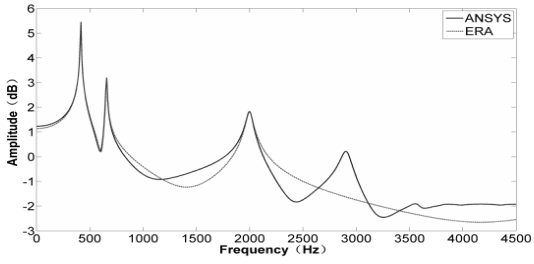


FIGURE 11. The FRF between disturbance and sensor 1

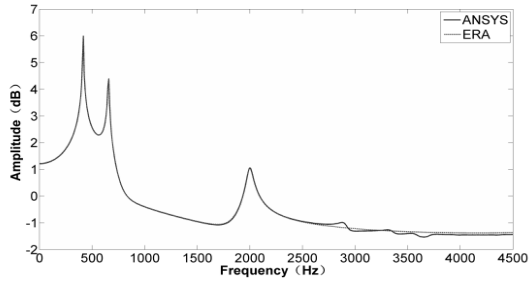


FIGURE 12. The FRF between disturbance and sensor 2

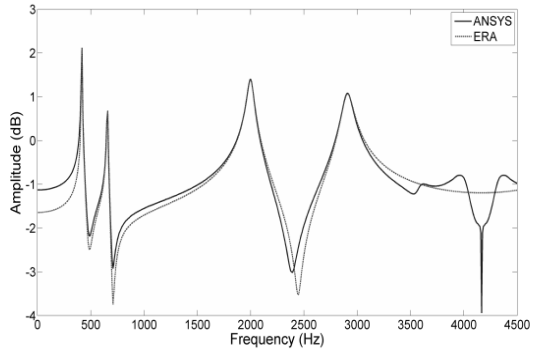


FIGURE 13. The FRF between actuator 1 and sensor 1

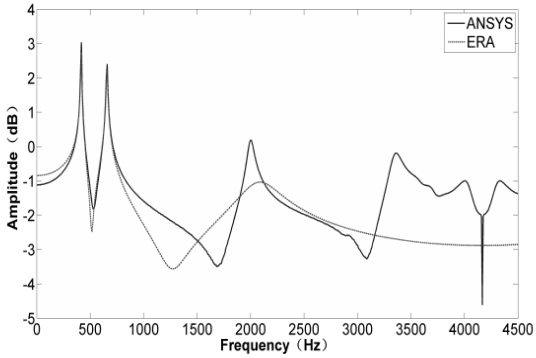


FIGURE 14. The FRF between actuator 2 and sensor 2

5. Active Controller Design and Simulation

5.1. H_2 CONTROLLER DESIGN

The state space equations (15) and (16) can be expanded by considering the disturbance being noise as following [14]:

$$x(i+1) = Ax(i) + B'u'(i) + Gw(i), \tag{17}$$

$$y(i) = Cx(i) + Du(i) + v(i), \tag{18}$$

$$z(i) = C'x(i) + D'u(i), \tag{19}$$

where $B' = [B_1 \ B_2 \ E]$, $u' = [u_1 \ u_2 \ 0]^T$, y and z are controlled and measured outputs, G is the input matrix of noises, C' and D' are measured output matrices, w and v are the process noise and measure noise. It is assumed that w and v are uncorrelated and have constant power spectral density matrices V_w and V_v. For the positive semi-definite matrix V_w, matrix G has following form

$$G = V_w^{1/2}. \tag{20}$$

In order to determine optimal gains the performance index J as (21) is minimized:

$$J^2 = E\left(\sum_{k=0}^{\infty} x^T(i)Qx(i) + u^T(i)Ru(i)\right). \tag{21}$$

In the above equation Q is a positive semi-definite state weight matrix and R is a positive definite matrix. So matrix C' can be defined through the weight Q

$$C' = Q^{1/2}. \tag{22}$$

Generally it is assumed $R = I$ and $V_v = I$, the minimum of J is achieved for the feedback with gain matrices as follows:

$$K_c = B^T S_{2c}, \tag{23}$$

$$K_e = S_{2e} C^t, \tag{24}$$

where S_{2c} and S_{2e} are solutions of the Riccati equations as follows:

$$S_{2c}A + A^T S_{2c} + C'^T C' - S_{2c}BB^T S_{2c} = 0, \tag{25}$$

$$S_{2e}A^T + AS_{2e} + GG^T - S_{2e}C^T CD_{2e} = 0. \tag{26}$$

The controller obtained includes a state estimator and an optimal feedback gain and the equations are given as follows:

$$\dot{\hat{x}}(i+1) = (A - BK_c - K_e C)\hat{x}(i) + K_e y, \tag{27}$$

$$u(i) = -K_c \hat{x}(i). \tag{28}$$

5.2. ACTIVE CONTROL SIMULATION

H_2 controller for DBA can be designed in the context of MATLAB environment, and then it can be incorporated with the FE model of DBA in ANSYS by using APDL

commands. The complete close-loop control system is shown as in Fig. 15.

An impulse disturbance is applied on the upper interface of model, and the displacement of the centre node of upper interface on Z-direction is measured. The results on open-loop condition and closed-loop condition are compared in Fig. 16. It is shown that the settle time is less than 0.06s by applying active control whereas it is more than 0.12s without control. Thus, active control can make system go into stable state much faster.

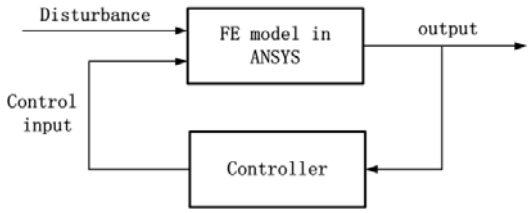


FIGURE 15. The scheme of active control in ANSYS

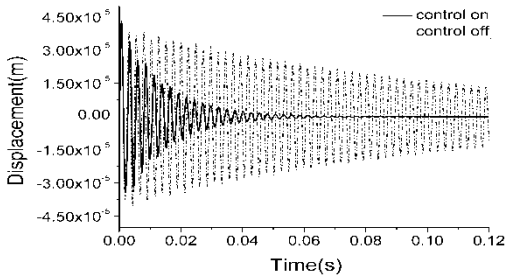


FIGURE 16. The displacement of centre node on upper interface

The voltages input to two actuators are shown in Figure 17 and 18. The maximum voltage amplitude on actuator 1 is 50V, while the maximum on actuator 2 is 80V. Both of them are within the PZT threshold.

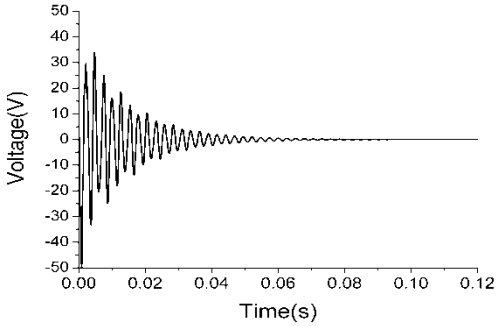


FIGURE 17. The voltage input to actuator 1

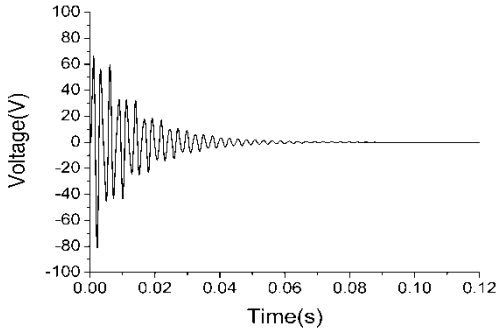


FIGURE 18. The voltage input to actuator 2

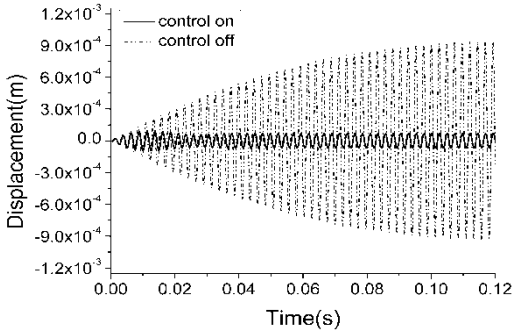


FIGURE 19. The displacement of centre node on upper interface

A sine disturbance with amplitude 10N and frequency f_1, which is the first natural frequency of DBA is applied on the upper interface. The displacements of the centre node of upper interface on Z-direction are compared in the cases of open-loop and closed-loop conditions. The results are shown in Fig. 19 and it is clear that the active control make the resonance being attenuated, and the response amplitude is decreased 20 dB than without control. The voltages input to two actuators are shown in Figure 20 and 21 respectively.

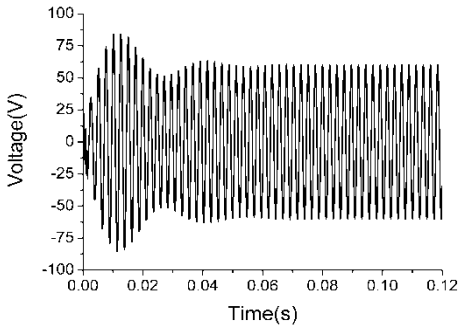


FIGURE 20. The voltage input to actuator 1

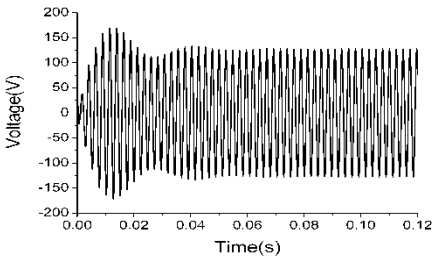


FIGURE 21. The voltage input to actuator 2

6. Conclusions

In this study, a DBA for active control is designed and tested by simulation in the context of ANSYS environment. The finite element model of DBA is developed in ANSYS, the folded beam and glue layers are meshed by SOLID45 elements and piezoelectric patches are meshed by coupling-field elements SOLID5. Control-system model is developed from the results of modal and transient analysis by employing the OKID and ERA methods, based on which H_2 controller are designed and incorporated with model developed in ANSYS by APDL commands. The closed-loop simulations of active control for DBA are performed under the impulse and sine excitation respectively, and the results indicate that DBA can attenuate vibrations effectively with low voltage as input. Even the validity of simulation in the context of ANSYS environment and the efficiency of identification algorithm have been approved and this simulation scheme works well for DBA in this paper, the experiment testing the real structure is still essential for practical implementation. Thus, in the future works more experiments are needed to check the dynamic and controlling characteristics of this actuator.

References

[1] Bandyopadhyay B., Manjunath T. C., Umapathy M. (2007) *Modelling, control and implementation of smart structures.* Berlin Heidelberg: Springer-Verlag
[2] Nader Jalili (2010) *Piezoelectric-Based Vibration Control*. New York: Springer Science Business Media, LLC
[3] Manning W. J., Plummer A. R., Levesley M C 2000 *Smart Mater. Struct* ,**9,** 932-9
[4] Gaudenzi P., Carbonaro R., Benzi E. (2000) *Composite Structures*, **50**(4), 373–9
[5] Singh S. P., Pruthi H. S., Agarwal V. P. (2003) *Journal of Sound and Vibration*, **262**(3), 563–75
[6] Xu S. X., Koko T. S. (2004) *Finite Elements in Analysis and Design*, **40**(3), 241–62
[7] Karagulle H., Malgaca L., Oktem H. F. (2004) *Smart Mater. Struct.*, **13,** 661–7
[8] Dong X. J., Meng G. (2006) *Journal of Sound and Vibration*, **297**(36), 680–93
[9] Wang Q., Wang C. M. (2000) *Smart Material and Structures*, **9**(4), 558–67
[10] Preumont A. (2002) *Vibration control of active structures: an introduction.* Kluwer academic publishers. 364 p.
[11] Wang Y. F., Cheng W. (2013) *Journal of vibration and shock*, **32**(22), 140-5
[12] *ANSYS Software* ANSYS Inc Canonsburg PA USA (www.ansys.com)
[13] Jer-Nan Juang, Minh Phan (1993) Identification of Observer/ Kalman Filter Markov Parameters: Theory and Experiments. *Journal of guidance, control, and dynamics*, **16**(2), 320-29, http://aero.tamu.edu/sites/default/files/jnj_symposium/jgcd93-16-2_OKID.pdf
[14] Wodek K., Gawronski (2004) *Advanced Structural Dynamics and Active Control of Structures*. New York: Springer-Verlag

Authors

Yun-Feng Wang, born 15.12.1983, Xingtai, China

University studies: Ph.D student School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics

Wei Cheng, 1961, Baoding, China

Current position, grades: Ph.D, professor in Beijing University of Aeronautics and Astronautics, did cooperative research in South Africa as a senior visiting scholar In 1999.
Scientific interest: structural dynamics

Using two-stage non-negative matrix factorization for topic recommendation in online social networks

Weihua Yuan[1, 2*], Hong Wang[1]

[1] *School of Information Science and Engineering, ShanDong Normal University, Jinan, Shandong province, China*

[2] *School of Computer Science and Technology, ShanDong Jianzhu University, Jinan, Shandong province, China*

Received 1 January 2014, www.tsi.lv

Abstract

This paper put forward a two-stage non-negative matrix factorization (TSNMF) for topic recommendation in online social network to solve the existing problems of mass data, extreme sparseness and cold start. In stage I, we use co-clustering to divide user-topic interest matrix into smaller sub-matrices called cluster-sub-matrices based on non-negative matrix factorization on interest-density matrix D. Each cluster-sub-matrix is much smaller than the original with similar internal interest pattern. In stage II, we use weighted non-negative matrix factorization algorithm to predict unknown items on each of cluster-sub-matrix directly. Experiments on real datasets show that TSNMF can not only gain high prediction accuracy on extreme sparse datasets, but avoid the problem of too much computation of NMF on the whole user-topic interest matrix, as well as the problem of the recommendation's quick local convergence.

Keywords: Recommender algorithm; NMF (non-negative matrix factorization); Clustering; Data sparseness

1. Introduction

Online social network is a kind of tool and platform for people to share views, opinions and experiences with each other through the Internet [1], which is a new application of network presence that provides users with great online participations. Social network is mainly divided into dating network, blog network, media-sharing network and instant communication network according to its functional properties. With the popularization of various forms of online social networks, a new one-way follow micro blog system as twitter and Sina Weibo is becoming a popular service, in which nodes are usually classified into users (subscriber) and subjects (publisher). Users as message subscribers on the one hand, build their one-way attention on topic nodes to form the one-way subscription relationship, while on the other hand, they would build two-directional concern of relationship with other users; on the contrary, subjects nodes as publishers, subscribed by a lot of users, achieve a lot more number of relationships of subscribers than that of the active attention and the two-directional concern. Main functions of recommendation system in online social network include recommendations of mutual concern relationship and subscriptions relationship. The former usually recommends users to a certain user whom he might want to follow in the near future or would be interest in through their common friends, contacts and mail lists, such as the recommendation of Tommy node to George; while the latter is about recommendation of subject nodes to users, like "air condition monitoring of Shandong province" to George, which would be achieved based on the users' interests [2].

This paper is mainly about how to recommend topics to users efficiently and effectively in such kind of one-follow relationship of online social networks.

2. Related works

At present, the recommendation algorithm is mainly divided into two categories: collaborative filtering based algorithm (CF) [5] and content based filtering algorithm (CBF) [3, 4]. The algorithm of CF would firstly compute the similarity of users or items, and then predict the unknown items based on the scoring of similar users or items. CBF algorithm usually recommends items to users according to correlation of contents of information and user profiles using technologies of information retrieval or information filtering. CBF usually could not flexibly combine useful information as much as possible (e.g. user interest), while CF relies too much on explicit or implicit rating data. As to the problem of sparseness existed in online social network, the algorithms mentioned above could not provide users with good recommendation effects, nor would they efficiently solve the problem of cold start (How to recommend items to new users and how to recommend new items to users). Therefore, being the new research focus, model based method was put forward [6], the essence of which is due to dimension

* Corresponding author. Tel: +86-15508676966;
E-mail:huahua_qingdao@126.com

reduction of data: determinants of user's rating on items can be attributed to some of the underlying elements with relatively fewer dimensions (such as film type, type of products etc.), besides, users within a type tend to rate similarly on goods of the same category. Therefore, model based methods need to approximate the original score matrix with a reduced and sparse decreasing probability distribution or low dimensional matrix. This kind of algorithm mainly includes clustering method, probability method and matrix factorization [6] and so on.

Matrix factorization algorithm (MF) generally decompose a high-dimensional matrix into the product of two or more low dimensional matrix to achieve dimension reduction, so as to study the properties of high dimensional data on a low dimensional space. Traditional MF algorithm mainly contains singular value decomposition (SVD) [9], non-negative matrix factorization (NMF) [10], and probabilistic matrix factorization (PMF) [11] etc. SVD algorithm allows decomposition with negative values, which is correct from the point of view of computation, however with no practical significance for many applications. The difference between NMF and SVD lies in the fact that the original non negative matrix will be approximated by the product of two non-negative sub-matrices. NMF is a sort of algorithm with local characteristic, whose expression is better than that of SVD, and the non-negative decomposition determines the intuitive physical meaning of its sub-matrices.

Recommendation problems of online social network, particularly for recommendations of subscription relations, are quite different from traditional systems. Firstly the problem of mass data -there are a large number of users and a variety of topics in online social networks, but recommendation algorithm needs to respond within the shortest possible time. Secondly lies in the problem of its extreme data sparseness. Mislove et al. [12] has pointed out that weak-relationship distribution features prominently in online social networks: very fewer users (topic nodes) own most of the connections (one-way subscription) while a great deal of users just have a small number of connections (one-way follow relationship like subscription). According to the statistics of 1000 sampling and 80 themes from the Sina weibo, the number of topics concerned by more than one half users is lower than 5% of the total. Consequently, collaborative filtering method based on binary relations could not achieve ideal recommendation effect when dealing with data with such extreme sparseness. The third is about the problem of dynamic data and cold start. Recommendation algorithm needs to react quickly and accurately because of the joining of new users or new topics, as well as the constantly changing of users' interests and concerns. However, content-based recommendation methods generally lack of enough diversity, which would make recommendation results converge quickly to a small scale set, thus losing more possible recommendations to users that they might be interested in.

Many researchers have put forward the method of clustering or dimensionality reduction to solve the above problem, among which includes models based on probability such as probabilistic latent semantic analysis (PLSA) [13], and models based on matrix factorization such as SVD, NMF, as well as models based on co-clustering method [14], which could make clustering from several dimensions simultaneously. However, algorithms mentioned above are still too computationally intensive and updated difficultly; as a result, they still could not solve these problems very well. Faced with the characteristics of data sparseness due to weak ties (one-way follow relationship) in online social networks, as well as the frequent update of subject information and user interests, we put forward a kind of dimensionality reduction prediction method: two-stage non-negative matrix factorization for topic recommendation (TSNMF). In stage I, users and topics to be recommended are clustered into some small sub-matrices called cluster-sub-matrices by means of co-clustering based on non-negative matrix factorization. Each cluster-sub-matrix after the cluster is much smaller than the original with similar internal interest pattern, which could predominantly lower the amount of calculation of stage Ⅱ. In stage Ⅱ, weighted non negative matrix factorization is utilized to fill and predict the blank elements of the smaller cluster-sub-matrix produced in stage I, to solve offline computation problem of the traditional dimension reduction method to achieve a more efficient subject recommendation in online social networks. Experiments show that TSNMF can obtain a more efficient subject recommendation in online social networks by reducing dimensions of the training data effectively and lowering offline computation remarkably, and at the same time maintaining higher prediction accuracy and resolving problem of sparseness in a better way.

The remainder of this paper is organized as follows. In Section 3, we provide mathematics description of topic recommendation system in online social networks and iterative process of non-negative matrix factorization. Section 4 presents our algorithm of TSNMF. Section 5 is the experimental result of our method on real datasets compared with other popular algorithms, followed by the conclusions and future work in section 6.

3. Topic recommendation systems in online social networks and the non-negative matrix factorization

3.1. DESCRIPTION OF TOPIC RECOMMENDATION SYSTEM IN ONLINE SOCIAL NETWORKS

Supposing there are a set of n users in online social network represented by a set $U = \{u_1, u_2, u_3, ..., u_n\}$, and each item in U is a row vector, called user-interest vector of user u_i and can be expressed as $u_i = \{a_{i1}, a_{i2}, a_{i3}, ..., a_{in}\}$. There are also a set of m

topics expressed as $T=\{t_1,t_2,t_3,...,t_m\}$, each t_j is a column vector, called top-rating vector, and the value of each element in t_j means its interest rating among different users.

Therefore all user-interest vector form a matrix of $n\times m$ called the user-topic interest matrix, expressed as $A=\{a_{ij}\}_{n\times m}$, each item $a_{ij}>0$ means that user u_i is interested in topic t_j, the value represents this user's attention rate or interest rate to the accordingly topic, and otherwise, 0.

In matrix A, the row vector shows a certain user's attention rate or interest rate of all the subjects and the column vector represents a certain subject's interest degree by different users. In this paper, we are to solve the problem of prediction of the unknown items in matrix A according to the existing users' interest rating of some topics, to recommend other subjects to users that they might be interested in or they would like to follow in the future.

3.2. NON-NEGATIVE MATRIX FACTORIZATION

The objective of our dimension reduction of original matrix A is to lower storage space and to save computing resources. Non-negative matrix factorization algorithm is a kind of multi-variable analysis method. Supposing matrix $A_{n\times m}$ with all elements negative, that is $a_{ij}\geq 0$, then its linear decomposition formula can be written as:

$$A_{n\times m}=B_{n\times r}C_{r\times m}+E_{n\times m},\ r<\min(n,m). \tag{1}$$

In which matrix B is called base matrix on behalf of users' interest dimension, while C is called coefficient matrix representing characteristic dimensions of the topic, and E is called approximate error, which can also be expressed as [15]:

$$E=A-BC. \tag{2}$$

In order to get close approximations of matrix A, $\|E\|$ is required to be as small as possible, and the algorithm is required to be of fast convergence. Therefore, the process of NMF is turned into the convergence method of the following formula:

$$\min_{B,C}\|A-BC\|. \tag{3}$$

Supposing noise obeys Gauss distribution:

$$p(A_{ij}|B,C)=\exp\left\{-\frac{1}{2}\left[\frac{A_{ij}-(BC)_{ij}}{\sigma_{ij}}\right]^2\right\}/\left(\sqrt{2\pi}\sigma_{ij}\right). \tag{4}$$

Among which σ_{ij} is the weight of each given observations, let

$$p(A|B,C)=\prod_{ij}p(A_{ij}|B,C). \tag{5}$$

The problem is transformed into the following solution of the maximum likelihood:

$$L(B,C)=\frac{1}{2}\sum_{ij}\left[A_{ij}-(BC)_{ij}\right]^2/\sigma_{ij}^2+\sum_{ij}\log\left(\sqrt{2\pi}\sigma_{ij}\right). \tag{6}$$

After the use of traditional gradient descent method, we get the following additive iterative formula:

$$\begin{aligned}B_{ik}&\leftarrow B_{ik}+\phi_{ik}\left[\left(AC^T\right)_{ik}-\left(BCC^T\right)_{ik}\right],\\ C_{kj}&\leftarrow C_{kj}+\phi_{kj}\left[\left(B^TA\right)_{kj}-\left(B^TBC\right)_{kj}\right].\end{aligned} \tag{7}$$

If we set $\phi_{ik}=\frac{B_{ik}}{\left(BCC^T\right)_{ik}}$, $\phi_{kj}=\frac{C_{kj}}{\left(B^TBC\right)_{kj}}$, then Equation (7) will be transformed into the following multiplicative iterative equation:

$$B_{ik}\leftarrow B_{ik}\frac{\left(AC^T\right)_{ik}}{\left(BCC^T\right)_{ik}},C_{kj}\leftarrow C_{kj}\frac{\left(B^TA\right)_{kj}}{\left(B^TBC\right)_{kj}}. \tag{8}$$

4. Two-phase non-negative matrix factorization algorithms

4.1. PROBLEM OF THE TRADITIONAL NMF

NMF algorithm could effectively reduce data dimensions, and decomposition result has obvious practical significance, in which sub-matrix B represents the dimensions of users' interest ratings on the topics, while C represents the feature dimensions of different topics. Yet the NMF algorithm has the following problems:

1) The high computational complexity and slow convergence, the complexity of each iterative is $O(\|B\|\times r+\|C\|\times r)$.

2) If you conduct matrix factorization on user-topic interest matrix A using NMF to finish the recommendations, the off-line calculation is much too computationally, and additionally, too many non similar

users or themes as interfering factors will lead to more inaccurate recommendations.

3) The locality of NMF leads to the problem that the decomposition results will depend on the ranks of matrix sequence of the original user-topic interest matrix A, and the result of the factorization usually converges to local optimum, rather than global optimal.

In order to reduce the complexity of online computations as well as to use more similar users and topic while rating the unknown items to exclude the interference of non related topic score information, the original matrix A needs to be pre-processed firstly. We make the work done through co-clustering based on NMF to narrow the size of data to be manipulated by NMF in stage II to enhance quality and density of the internal rating data in each category of the cluster.

4.2. STAGE I: CO-CLUSTERING BASED ON NMF

Clustering of the original matrix is a kind of method widely used in dimension reduction, because with the unceasing expansion of scales of online social networks, the number of users will increase accordingly, which will render it increasingly more time-consuming to find the nearest neighbours. As a result, it is very necessary to assign users and topics into different bureau categories based the similarity to exclude the non-related factors to enhance the efficiency of online neighbour search. We finish dimension reduction using clustering to lose information as little as possible. Firstly, we define the core user set U_{den} and the interest-density matrix D.

4.2.1. Core user set U_{den} and interest-density matrix

The interest density of each user u_i is defined as the proportion of non-zero items of topic-interest row vector u_i, expressed as $den\_in(u_i)$. A core user is defined such that $den\_in(u_i)$ of user u_i surpasses a threshold λ (usually a value of 8%-10%), and the set of core users can be represented as the following:

$$U_{den} = \{u_i | u_i \in U, den\_in(u_i) > \lambda\}. \tag{9}$$

The vectors of core users U_{den} from user-topic matrix A compose the interest-density matrix D. We conduct non-negative matrix factorization on matrix D to get core clustering subsets of D, and then we finish clustering on the original matrix A. The following is Algorithm 1.

4.2.2. Algorithm of co-clustering based on NMF of Stage I

Algorithm 1 of Co-clustering based on NMF is as follows:

Input: Interest-density matrix D, user-subject interest matrix A.

Output: User-topic interest subsets of each category after clustering (called cluster-sub-matrix).

Step 1: The building of Core user set U_{den} and interest-density matrix D.

Step 2: The process of NMF on U_{den}, carrying out the multiplicative iterative process according to Equation (8) until the convergence of $\min_{B,C} \|D - BC\|$.

Step 3: Co-clustering analysis of sub-matrix B and C, if $j = \arg\max_j b_{ij}$, then user i is assigned to subject j. After clustering of D, the various core cluster subsets are got.

Step 4: Taking the interest-density clusters as centre to finish co-clustering on non dense sub-matrix $A - D$ to get $cluster - sub - matrix$ of each category, which of each has similar interest pattern inside.

4.3. STAGE II: WEIGHTED NON-NEGATIVE MATRIX FACTORIZATION ON $cluster - sub - matrix$ OF STAGE I, TO FILL UNKNOWN ITEMS TO COMPLETE RECOMMENDATIONS

4.3.1. Weighted non negative matrix factorization

According the above, the user-topic interest matrix is very sparse in online social networks. So in order to approximate the original matrix, there will be a lot of zero elements being added to new low-dimensional matrices if the zero items of the original matrix are not operated, which does not conform to our recommendation requirements to predict the ratings of these unknowns. So when we carry out non-negative matrix factorization on the $cluster - sub - matrix$, we need to adopt weighted NMF algorithm to omit the corresponding zero items in calculation errors. The goal of calculation is to minimize the loss function below [15]:

$$\min \left\| I_{n\times m} \otimes \left(A_{n\times m} - B_{n\times r} C_{r\times m} \right) \right\|_F^2, \tag{10}$$

s.t.

$$B_{ij} \geq 0, C_{ij} \geq 0, r < \min(n, m), \tag{11}$$

Among which matrix I is a label matrix with its elements defined as:

$$\begin{cases} I_{ij} = 0, & if\ A_{ij} = 0 \\ I_{ij} = 1, & if\ A_{ij} > 0 \end{cases}, \text{I} \tag{12}$$

The according multiplicative iterative formula is transformed into Equation (13) and (14):

$$B_{ij} = B_{ij} \frac{\left((I \otimes A)C^T\right)_{ij}}{\left((I \otimes (BC))C^T\right)_{ij}}, \quad (13)$$

$$C_{ij} = C_{ij} \frac{\left(B^T (I \otimes A)\right)_{ij}}{\left(B^T (I \otimes (BC))\right)_{ij}}, \quad (14)$$

Among which $\otimes$ represents element-wise multiplication of the matrix.

4.3.2. Algorithm of stage II

Input: $cluster-sub-matrix$ after clustering of stage I, user-topic interest matrix A, and dimension limit r.

Output: User-topic interest matrix with unknown items filled with predicting ratings.

Step1: Random initialization of B and C.

Step2: The adjustment of elements in both matrix B and C, by performing multiplicative iterative process according to Equation (13) and Equation (14).

Step3: Repeating Step2, until NMF algorithm converges.

Step4: The implementation of subject recommendation in online social network according to sub-matrix B and C.

4.4. COMPLEXITY ANALYSIS

Time complexity of stage I is $O(number(iter) \times \|D\| \times r_1)$, time complexity of stage II is $O(number(iter) \times size(cluster-sub-matrix) \times r_2)$, in which $number(iter)$ is the number of iterations, and $size(cluster-sub-matrix)$ is the number of non-zero elements in each $cluster-sub-matrix$. The time complexity of the prediction phase is constant level so the system can be updated in real time due to the limitation of rating scale of each category.

5. Experimental results and the analysis

5.1. EXPERIMENTAL DATA

In this paper our experimental datasets were crawling through an open system of a certain micro-blog platform using their API, the actual running data was from March 2013 to July 2013. The original dataset consists of about hundreds of thousands of users, more than 300 themes and more than 300 million pieces of micro-blog information. During the experiment, we firstly conduct pre-processing to the original dataset using co-clustering based on NMF, to transform it into many different $cluster-sub-matrices$, the final results of the experiment is the average values of these various experimental data. Hardware configuration of this experiment is a Hadoop cluster system of 3 nodes, each of which is a PC machine with Core™ i3-2120, CPU@3.30GHZ, and memory 2GB, and the operation system is Linux (Ubuntu 12.04). The system was implemented by Python and Java.

5.2. EVALUATION INDEX

The accuracy of predicting ratings is expressed as closeness between a user's real rating and the prediction score of the algorithm. We use RMSE (root mean square error) as the standard of predicting accuracy, that is, to ratings not appearing in training set but in test set we will compute their RMSE.

$$RMSE = \sqrt{\frac{1}{|TEST|} \sum_{(i,j,R_{i,j}) \in TEST} \left(R_{i,j} - estimated\right)^2}, \quad (15)$$

In which TEST represents the test set, and $|TEST|$ is the number of people's ratings in test set. The smaller value of RMSE, the higher accuracy of the predicting algorithm is.

The reference algorithms with our TSNMF include:

1) Collaborative filtering recommendation algorithm CF based on Top-K similarity, which uses the Pearson coefficient to compute similarity of users, and the final recommendation results is according to Top-K.

2) Non-negative matrix factorization (NMF) algorithm, which carry matrix factorization on the original user-topic interest matrix A directly to realize top recommendations.

The above two algorithms were realized based on the open source machine learning library-Apache Mahout.

5.3. IMPACTS OF DIMENSION r ON RUNNING TIME OF TSNMF

The first experiment we designed is to verify impacts of implicit dimension r on execution time of TSNMF. The results is shown in Figure 1, from which we can see that with the increase of r, the running time of TSNMF algorithm becomes increasingly long and the operating efficiency really declines substantially. However, we could also see that when the value of r is between 10 and 25, the running time of the system remained approximately stable. Consequently, we can draw a conclusion that in order to enable the running efficiency as well as ensuring the recommendation accuracy requirements simultaneously, the value of r needs to be limited to [10, 25].

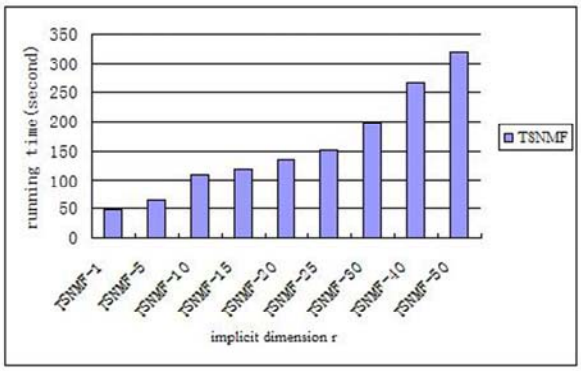


FIGURE 1. Impacts of dimension r on running time of TSNMF

5.4. COMPARISON OF RECOMMENDATION PRECISION BETWEEN SVD, NMF AND TSNMF

The second experiment we designed is to verify different recommendation precision on the three algorithms SVD, NFM and TPNFM, and the result is shown in figure 2. It is quite clear that our TPNFM has the highest accuracy compared to SVD and NMF, which is due to the efficient co-clustering of stage I, as well as the weighted NMF running on $cluster-sub-matrices$. The experiment proved that our TSNMF did increase and provide high top recommendation accuracy on online social networks.

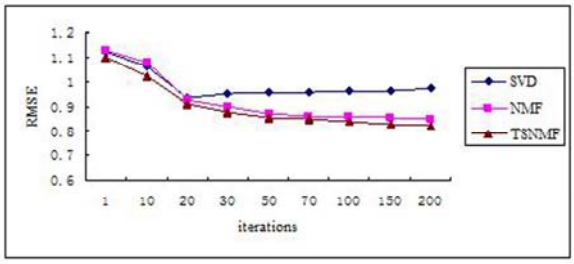


FIGURE 2. Comparison of recommendation precision between SVD, NMF and TSNMF

5.5. IMPACTS OF DATA SPARSENESS AND PARAMETER r ON TOPIC RECOMMENDATION ACCURACY

Drawing on the experience of literature [16], we chose one part of the experimental data randomly from experimental data as training dataset, and the rest as test dataset. The proportion of the training data is 80%, 50%, 20% respectively. The purpose of this experiment is to verify the effect of the algorithm under different data sparseness. The result is shown from Figure 3 to Figure 5.

Figure 3 is the RMSE variation tendency of algorithm SVD with 80%, 50%, 20% as training dataset respectively; Figure 4 and Figure 5 are about that of NMF and TSNMF, respectively.

From Figure 3, we can see that the RMSE of SVD begins to increase rapidly with the dimension r rising over 25 and 20% as the training data, and the same as Figure 4 of NMF, with a relatively slower growth trend.

Figure 5 shows that the RMSE of our TSNMF keeps the same decline trend, although RMSE with 20% as training data is higher than that of 50% and 80% as training data respectively, which shows that our TSNMF provides a better way to resolve problem of data sparseness.

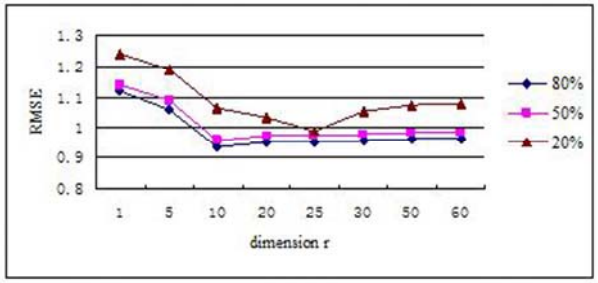


FIGURE 3. SVD algorithm 80%, 50% and 20% as the training data respectively

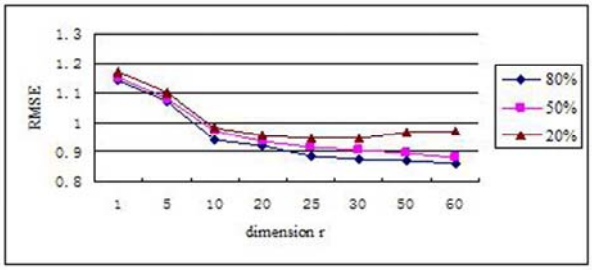


FIGURE 4. NMF algorithm 80%, 50% and 20% as the training data respectively

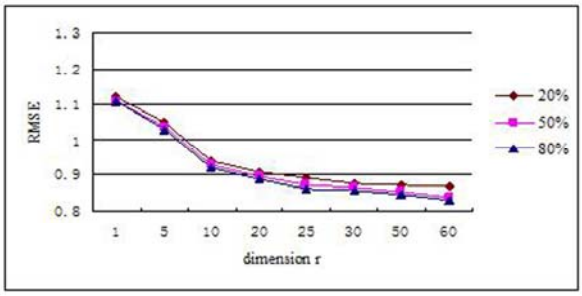


FIGURE 5. NMF algorithm 80%, 50% and 20% as the training data respectively

6. Summary and prospect

In this paper, we put forward a kind of two-stage non-negative matrix factorization TSNMF to provide topic recommendation of predicting ratings in online social networks. In stage I we form the interest-density matrix D by extracting core user interest vector $den\_in(u_i)$, and then we finish co-clustering of user-topic interest matrix A after non-negative matrix factorization of D, and in stage II we realize rating prediction of topics by means of weighted NMF algorithm to each $cluster-sub-matrix$. The experiments show that our algorithm TSNMF can gain relatively high accuracy on extreme sparse datasets, and can also provide with recommendations of rich diversity in the context of cold start, which has not only avoided the problem of computing excessively on user-topic interest matrix but also eliminating the problem of quick and local convergence of the result.

In future work, we hope to introduce a probabilistic model in the stage of co-clustering to make the sub-clusters more flexibly because the present clustering operation is mainly based on the sparse user-topic interest matrix which might not provide accurate clustering, and further more we also hope to introduce a feedback mechanism to realize feedback based on user's real interest on recommendation results, to realize dynamic optimization of the algorithm to obtain high efficiency.

Acknowledgements

This work is supported by the Technology Program of Shandong Province under Grant No. 2012GGB01058, China; A Project of Higher Educational Science and Technology Program of Shandong Province (J12LN31), China; A project of Soft Science Research Project of Shandong Province (2013RKB01310).

References

[1] Kaplan, Andreas M., Haenlein M. (2010) *Business Horizons,* **53**(1), 59-68
[2] Chen K.-H., Han P.-P., Wu J. (2013) User clustering based on social network recommendation. *Chinese Journal of Computers,* **2,** 349-59
[3] Belkin N., Croft B. (1992) *Communications of the ACM* **35**(12), 29-38
[4] Balabanovic M., Shoham Y. (1997) *Communications of the ACM,* **40**(3), 66-72
[5] Deshpande M., Karypis G. (2004) *ACM Trans on Information Systems,* **22**(1), 143-77
[6] Marlin B. (2004) Collaborative Filtering A machine learning perspective *MS.Thesis.* Toronto: University of Toronto.
[7] Lee D. D., Seung H. S. (1999) *Nature,* **401**(6755), 788-91
[8] Hofmann T. (2004) *ACM Transactions on Information Systems,* **22**(1), 89-115
[9] Golub G., Kahan K. (1965) Calculating the singular values and pseudo-inverse of a matrix, *Journal of the Society for Industrial and Applied Mathematics*, **2**(2), 205-24
[10] Lee D. D., Seung H. S. (2000) Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems,* 556-62
[11] Salakhutdinov R., Mnih A. (2008) Probabilistic matrix factorization. *Advances in Neural Information Processing Systems,* **20**(3), 432-51
[12] Mislove A., Marcon M., Gummadi Krishna P., Druschel P., et al. (2007) Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. San Diego, CA, USA, 29-42
[13] Blei D. M., Ng A. Y, Jordan M. I. (2003) Latent Dirichlet allocation Journal of Machine Learning Research, **3**(3), 993-1022
[14] Shah H. H., Banerjee (2007) A Bayesian co-clustering. In *Altman R. ed. Proc. of the ICDM 2008.* Washington: IEEE Computer Society Press, 530-9
[15] Sajda P., Du S., Parra L., et al (2003) Recovery of constituent spectra using non-negative matrix factorization. *Proc SPIE*, 5207, 321-31
[16] Liu N. N., Yang Q. (2008) Eigenrank: A ranking-oriented approach to collaborative filtering. *Proc. of the SIGIR 2008*, 83-90 http://www.cs.ust.hk/~qyang/Docs/2008/SIGIR297-liu.pdf
[17] Hong Wang (2012) Bi-direction Link Prediction in Dynamic Multi-dimension. *Networks Journal of software (supplement),* **11,** 176-85
[18] Wang Hong, Yuan Wei Hua, Zhou Qian (2013) Backward Direction Link Prediction in Multi-relation Systems ITME2013. Xining China, p. 2031-36
[19] Yuan Wei-Hua, Wang Hong, Fan Zhong-Yong (2013) Research on Optimization of Resources Allocation in Cloud Computing Based on Structure Supportiveness. *ITME2013.* Xining China, 849-58

Authors

Weihua Yuan, 1977, Qingdao, Shandong, China

Current position: lecturer
University studies: a PhD student working under supervision of Dr.Hong Wang, Shandong Normal University
Scientific interest: complex networks, data mining and big data cloud
Publications: about 10
Experience: She gained Master's in management science and engineering from Shandong normal university in 2000, and Bachelor's degree in computer application from Guizhou university in 2006. She has been working in school of computer science and technology, Shandong Jianzhu University since 2006, and she is also a PhD student working under supervision of Dr.Hong Wang in Shandong Normal University since 2012

Hong Wang, 1966, Tianji, China

Current position: Professor
University studies: she gained Master's and Bachelor's degree in computer software of Tianjin University in 1988, gained PHD of the Chinese academy of sciences in 2002.
Scientific interest: complex networks, workflow, mobile agent, social software and big data cloud
Publications: over 60
Experience: she gained Master's and Bachelor's degree in computer software of Tianjin University in 1988, gained PHD of the Chinese academy of sciences in 2002. She has been working in Shandong Normal University since 1991. Currently she is a professor and doctoral supervisor of Computer Science department at School of Information Science and Engineering. She writes and publishes over 60 academic articles

Optimization of medical information systems by using additional factors

D. Zagulova[2], R. Muhamedyev[1*], I. Ualiyeva[1], A. Mansharipova[3], E. Muhamedyeva[4]

[1] *International IT University, CT&SE department, Jandosova/Manasa 34a/8a, Almaty 050040, Kazakhstan*

[2] *Tomsk State University, Lenin Prospekt 36, Tomsk, 634050, Russia*

[3] *Kazakh-Russian Medical University, Torekulova str. 71, Almaty 050004, Kazakhstan*

[4] *Riga Technical University, Kalku 1, Rīga LV-1658, Latvia*

Received 1 January 2014, www.tsi.lv

Abstract

Increasing longevity is one of the most important problems of modern Gerontology. Solution of these problems is connected principally with the use of information and communication technologies. Creation of a comprehensive health information system requires consideration of many factors, such as qualitative screening system based on patients' self-assessment, identification of possible errors that affect decision-making and patients' personal characteristics. The work presents the results of elderly Almaty and Almaty Region population survey conducted with the help of Active Longevity Portal designed for data collection, analysis and assistance to the elderly population of Kazakhstan. The results showed that the number of medical consultations is directly related to health self-assessment and anxiety levels. Detection of cardiovascular diseases (CVD) with the help of effort angina self-assessment demonstrated low sensitivity. Correlation between the Kettle's index of effort angina self-assessment, the impact of Physical Component Score (PCS) of SF-12 test onto the manifestation of cardiovascular disease in hereditary background, anxiety level and coronary heart disease manifestation, impact of Health Survey estimated by Physical Component Score (PCS) and Mental Component Score (MCS) SF-12 test onto the correspondence between Effort Angina Questionnaire and CVD patient state was detected. Studies showed that detection of diseases through Questionnaire Survey self-assessment in certain situations may lead to significant errors. Consideration of these factors will help to build a more powerful information system in which personal data will be combined with clinical data and expert estimates.

Keywords: gerontology, Kazakhstan population, information technology, cardiovascular diseases, medical information system

1. Introduction: *Actuality of the Research*

Gerontological research is associated with the rapid aging of the population in developed countries. This process is caused not only by the birth rate decrease, but also by a life expectancy increase. For example, according to European experts in 2060 one third of European population will reach age 80 or elder [1]. The same processes occur in Japan [2]. Researchers predict population decrease in China, India, Brazil, Russia and increase of older people number [3].

Kazakhstan is a developing country, but its population is also aging. In Kazakhstan, as of January 1, 2010 the number of people aged 65 years and older was 7.14% of the total population of the country. UN experts consider Kazakhstan to be the state with accelerated aging. By 2050 25% of elder people are expected in the country according to forecasts.

Because of this, Kazakh society faces the problem of active longevity prolongation reducing the costs of health care for elder people and increasing their demand in the labour market in old age.

Improving care for elder people, comprehensive solution of their medical-biological, social, and psychological aspects is one of the priorities defined by the State program "Salamatty Kazakhstan" for 2011-2015, approved by the Decree № 1113 of the President of the Republic of Kazakhstan d.d. 29.11.2010.

Economic analysis shows that it is necessary to find new methods and technologies, which will help to improve the quality of service and to reduce costs. Initiatives that are offered in different regions are usually connected with the use of information and communication technologies (ICT).

At the same time we have to point out that, the potential of ICT in Gerontological market is very high: the fact is that Europeans aged 65+ own funds over € 3,000 milliard [4]. In Kazakhstan, the situation is different, but due to the growth of elderly population (1,634,974 people by 2011), Internet usage increase, income growing, as well as increase of the number of

* Corresponding author. Tel: +7 707 684 43 70;
E-mail: ravil.muhamedyev@gmail.com

specialists in the field of Gerontology and Geriatrics the potential of ICT will also be high.

With the introduction of principles of pharmaco-economic analysis and health-economic standards into the work evaluation of health care organizations, not only a correct statement of the clinical diagnosis and the reasoned purpose of treatment, but also prevention and prediction of diseases became very important in everyday medical practice.

An adequate solution to these problems is possible while optimizing Medical Information Systems (MIS). MIS, are the most important component of accounting and control of medical practice, they allow to maintain an electronic version of the medical record, plan research, obtain diagnostic information and provide data for scientific research.

In the medicine of the future, a pivotal role belongs not to the treatment of diseases but to their prevention and early prediction. With adequate prevention and prediction of diseases, it is possible not only to reduce morbidity, but also improve the economic performance of the healthcare system.

Despite the advances in modern medicine, the percentage of errors in medical practice remains high [5]. One of the purposes of computer systems is to support decision making to reduce the risk of erroneous diagnoses and treatment. In this case, Electronic health record (EHR) is the base, which contains comprehensive data for the analysis and development of medical science and practice. However, there are many challenges to be faced in the development of effective and adequate information system.

First of all it is necessary to standardize the basic set of data that will provide comprehensive health assessment, risk assessment and disease diagnosis, and will allow optimizing case management and solving problems of organization and control.

Cost-effectiveness of health is affected by many factors including frequency of patient address to qualified medical professionals and duration of treatment of diseases, which in turn depends on the accuracy of diagnosis and case management.

Multifunctional portal aimed to optimize the health care of the elderly population is being developed as part of the scientific and technical program of the Ministry of Health of the Republic of Kazakhstan [6].

The ultimate goal of the portal development is the creation of an information resource that will perform the following functions:

Function of questionnaire self-entry by patients or their relatives.

Function to gather statistical information on the questionnaires from the portal database.

Analytics function.

Self-diagnosis function.

Function of economic analysis and recommendations generation.

To implement the MIS first of all it is necessary to define the parameters which will be recorded in patients and will prove sufficient for the planned analytical and forecasting aims. This publication presents the first stage of the required criteria selection. Some issues related to the influence of personal factors on case management processes, which should be taken into account while designing both medical information system (MIS) and the EHR, are considered. The work consists of the following parts:

In Part 2 the basic functionality of the Active Longevity Portal as part of the MIS is discussed.

In Part 3 the field of study and the methods are considered.

In Part 4 the results of the research are presented.

In Part 5 the necessity of both clinical and other personal characteristics of the patients in the MIS design and the impact of the analysis on the development of the Active Longevity Portal is discussed.

In conclusion the results of the analysis and their impact on the development of MIS are summarized.

2. Objectives and Functions of the Portal

The portal is supposed to be an input point of "Gerontological space" of Kazakhstan including information about major aspects of Kazakhstan population aging. It will be a platform for experience exchange, publication of research results and reception of generalized and personified information about elder people health (Fig. 1).

Based on the collected information using methods of scientific forecasting, searching patterns in the data and forecasting in processes of aging by experts in gerontology and geriatrics, expert systems on various aspects of aging process will be developed.

Portal users at this stage:

Doctors (managers for interviewers).

Interviewers (persons who gather and enter questionnaires).

Managers.

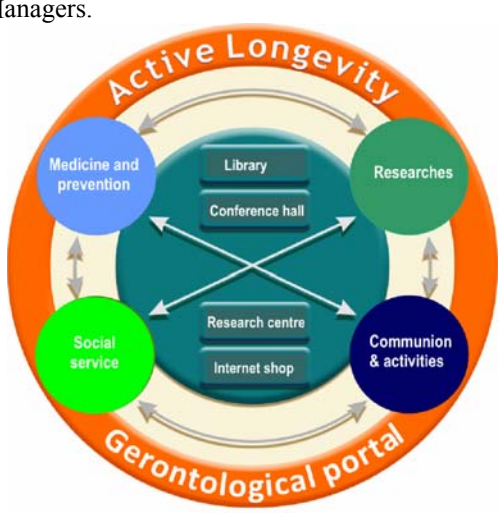


FIGURE 1. Functional interconnection of Active Longevity Portal

Portal functions at this stage are the following:

1. Questionnaires selection according to various criteria.
2. Questionnaires input, editing and deleting.
3. Retrieval of information about interviewers.
4. Gathering of statistics from questionnaires entered into the data base.

Prospective users of the portal in the future (additional to initial portal users):

- Specialists in the gerontology.
- Any other medical specialists, practitioners, interns, residents and graduate students.
- Health facilities.
- Sanatorium and rehabilitation centres.
- Social services.
- Elder citizens, their relatives and carers.
- Organizations that use volunteer work.

Any interested organizations and individuals, including entrepreneurs, merchants, sponsors and patrons, etc.

Health insurance

Medical equipment and medicine manufacturers and distributors.

Possible functions of the portal in the future:

- Function of questionnaire self-entry by patients or their relatives.
- Function to gather statistical information on the questionnaires from the portal database.
- Analytics function.
- Self-diagnosis function.
- Placement of scientific publications and scientific journal of the medical community.
- Relation: patient - portal - clinic. Information exchange between the patient and the portal, the patient and the clinic, the clinic and the portal.
- Portal users forum on Gerontology or other medical issues.

3. Methods

This work contains analysis of the data obtained in the course of medical screening tests of residents of Almaty city and Almaty region (the Republic of Kazakhstan), which was conducted by the Active Longevity Portal. The study addressed two sets of problems. Firstly, the problems of healthcare on the medical examination of the population were considered. Secondly, the problem of designing a screening Internet-system and creation of a medical database and testing it on real data were solved. In total 3,032 people including 1,660 women and 1,372 men aged 36-94 were examined.

All data were entered into the developed information system and then exported for analysis in MS Excel and Statistica 8.0. The data were collected through the portal system Active Longevity.

To determine anxiety and depression there was used a self-assessment scale "Hospital Anxiety and Depression Scale" (HADS), which is designed for a screening identification of anxiety and depression [7]. Depending on the result of HADS all patients were divided into 3 groups.

To detect coronary heart disease (CHD) in population mass screening it is necessary to use standardized and reproducible methods. In this study a standard questionnaire of WHO which is widely used in the practice of population-based surveys was used to identify the effort angina [8].

Hereditary background was assessed in patient survey, recording diseases of their parents.

In Statistica 8.0 for multiple comparisons using ANOVA method there was used Tukey HSD for unequal N and Scheffe test, a rough criterion which is suitable for cases where there is a suspicion of inequality of variances between the samples.

To compare two proportions there was used Statistica 8.0, p-level is computed based on the t-value for the respective comparison.

4. Findings

4.1. THE INFLUENCE OF PATIENTS' PERSONALITY FACTORS ON THE ACCURACY OF CARDIOVASCULAR DISEASE (CVD) RISK ASSESSMENT

EHRs are focused on the automation of the treatment process, including the optimization of diagnosis and prediction.

A large number of publications indicate the influence of patients' personality factors on course and treatment of various diseases.

However, Health Organizations and EHR in particular, are frequently not directed to use patients' personality factors in the diagnosis and treatment. In the bibliography there are only a few reports on its impact and the need for inclusion of additional indicators that are better and can describe the patients' condition more systematically in the EHR standard [9, 10, 11].

Cardiovascular diseases are the most common indication for hospitalization among adults over 65 years [12]. CVD is a very common disease, one of the leading causes of death, as well as temporary and permanent disability of the population in the developed countries. According to WHO statistics the largest proportion of deaths is caused by cardiovascular disease (48%) [13]. WHO recommends to conduct periodic health examinations to ensure timely detection of risk CVD. Screening is a common strategy in many countries to reduce the disease rate through early detection and intervention [14].

The fate of patients depends not only on the adequacy of the treatment, but also on the quality and timeliness of

diagnosis of the clinical forms of the disease, which require the emergent help or urgent hospitalization.

In preventive medicine for large-scale identification of individuals at risk of coronary heart disease questionnaire is often used. In the present study the hypothesis about the impact of some personality factors such as anxiety, Physical Component Score (PCS) and Mental Component Score (MCS) (SF-12) on questionnaire identifying the risks of coronary heart disease (CHD) is tested.

Increased anxiety indicates the patient's propensity to perceive an extensive range of life situations as threatening and respond with a large degree of anxiety. Anxiety disorders occur with 5-10% of the adult population and, for example, in case of myocardial infarction it happens 2-3 fold often [15].

The essential problem in identifying CVD can be low sensitivity of questionnaires used in screening. In the group under study 65.1% of the patients with CVD questionnaire revealed no effort angina. When the respondents were split into 3 groups with different levels of anxiety, it was found out that:

- with a low level of anxiety questionnaire method showed no effort angina in 83.1% of patients with CVD;
- with a high level of anxiety angina was not diagnosed in 43.3% of patients with CVD.

The results show that self-assessment of effort angina may not detect CVD of individuals with a normal level of anxiety. Z. M. Lohanova discloses [16] possible reasons for incorrect diagnosis of angina, which may be associated both with the patient's personality and inefficiency of diagnostic procedures.

In addition, the work states that the identification of symptoms and assessment of the severity of the disease may depend on the method of collecting information that is on self assessment or medical documents [17]. In particular, self assessment in the evaluation of cardiovascular system state proved to be less effective.

This paper presents the research of the influence of Health Survey, estimated with Physical Component Score (PCS) and Mental Component Score (MCS) of SF-12 test, on the correspondence between the Questionnaire assessment of effort angina and patient CVD.

Analysis of the data showed that:

- at the high level of Physical Component Score questionnaire method showed no effort angina in 78.6% of patients with CVD;
- at the low level of PCS effort angina was diagnosed in 31.0% of patients with CVD.

Thus, with higher rates of Physical Component Score, as well as with self-anxiety assessment case self-assessment of effort angina cannot detect CVD in patients having it.

At the same time, no influence of Mental Component Score SF-12 on error rate was detected when using the questionnaire method for effort angina determination.

The research showed that the error rate with effort angina questionnaire estimation method may depend on Kettle's index (weight-height index), which is calculated as weight in kilograms by height in meters. Kettle's index allows to assess physical development and take into consideration body parameters of individuals (related to health, health care, diet, etc.).

Data analysis showed that 83.9% of the normal Kettle's index, as in the case of Physical Component Score, self gina can not detect the presence of CVD in patients with this disease.

4.2. SIGNS OF HEREDITARY TAINTED CVD

Hereditary background as well as identification of factors, which influence the development of hereditary tainted diseases are of great importance in disease prediction. The relationship between the level of anxiety and depression and the frequency of manifestations of hereditary tainted cardiovascular disease was revealed.

The study showed that:

- in the group of individuals who have parents with cardiovascular disease, the frequency of such diseases statistically significant ($p < 0,01$) was different in the groups with low, medium and high anxiety and was 52.9%, 77.5% and 100.0% respectively;
- in the group of individuals who have parents with cardiovascular disease, the frequency of manifestations of cardiovascular disease statistically significant ($p < 0,01$) was different between the groups with high and low depression and was 60.8% and 100.0% respectively.

The influence of Physical Component Score (PCS) of SF-12 test on manifestation of cardiovascular disease with hereditary background was revealed in the present work. In the group of individuals who have parents with cardiovascular disease, the frequency of manifestations of cardiovascular disease statistically significant ($p < 0,01$) was different in the groups with low and high PCS and was 80.0% and 34.8% respectively.

Observed relationship may be caused by various reasons, which will not be discussed in this article. In the perspective of this study, it is important that some of the personal characteristics of patients can significantly affect the appearance and course of various diseases, which, in turn, must be considered when designing medical information systems.

4.3. INTERCONNECTION OF MEDICAL CONSULTATIONS FREQUENCY AND PATIENTS' ANXIETY

The effectiveness of MIS effects the quality of healthcare, which, in turn, depends on the success of the solution of organizational and managerial tasks. Among

these tasks, the most important are the processes of optimization of medical work, which are based on forecasts of the medical consultations frequency, as well as prognosis of the number of medical staff with required qualifications and skills.

The research of the scientific sources shows that the behaviour of individuals depends on personal factors and cannot be unambiguously predicted for the entire population. Personal factors include features of the individual, which are not part of health changes or health indicators.

They are gender, race, age, social background, education, occupation, lifestyle, life experience, psychological characteristics, of which some or all may affect health, diseases and treatment of individuals. However, at present, organizational and managerial task are solved without taking into account possible influence of personality characteristics of patients. For example, in case of hypochondria, which can be a symptom of depression, patients pay too much attention to their health and have frequent complaints of malaise.

The hypothesis on the impact of the level of anxiety and depression on the frequency of patient's medical consultations was tested in the present study. Analysis of the data showed that the number of medical consultations is interconnected with the level of self-assessed health and disease rate and with the level of anxiety. It should be noted that the relationship between the number of medical consultations and anxiety levels was not observed in 55 year old patients and younger.

The following relationship was not different with men and women, so the graphs represent men and women over 55. Tukey HSD for unequal N and Scheffe test showed a statistically significant difference ($p < 0,01$) levels of self-assessed health and the frequency of medical consultations among all groups with different levels of anxiety (Fig. 1, Fig. 2).

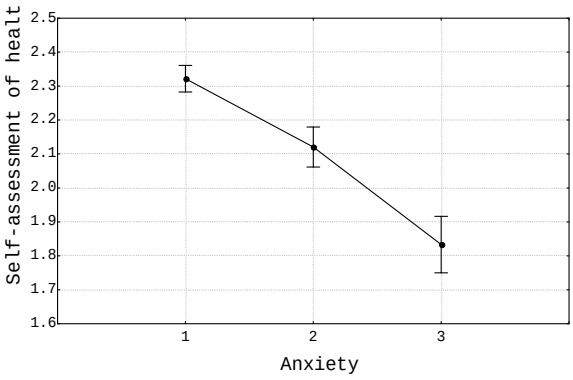


FIGURE 1. Graph of means and confidence interval (95.00%): levels of self-assessed health at different levels of anxiety

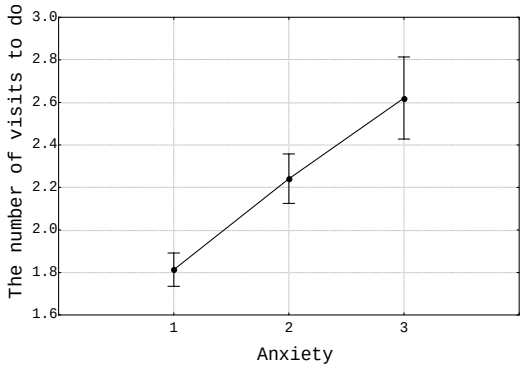


FIGURE 2. Graph of means and confidence interval (95.00%): number of medical consultations at different levels of anxiety

Number of medical consultations was statistically significantly higher at the lowest level of health self-assessment (Fig. 3).

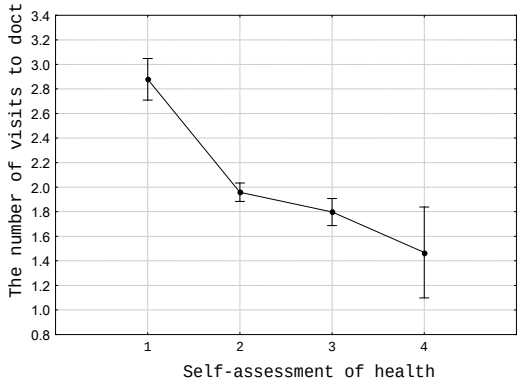


FIGURE 3. Graph of means and confidence interval (95.00%): number of medical consultations at different levels of self-assessment of health

Based on the data below, to exclude the effect of the level of health, the relationship of anxiety and frequency of medical consultations studied in groups with the same level of self-rated health.

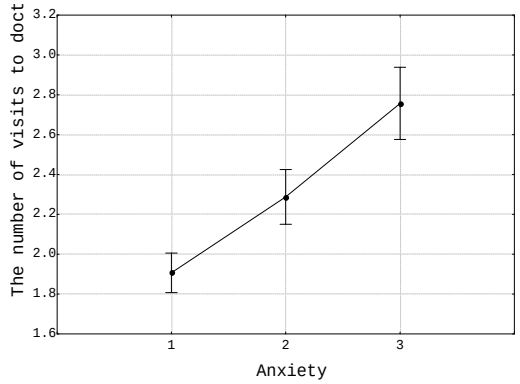


FIGURE 4. Graph of means and confidence interval (95.00%): number of medical consultation at different levels of anxiety in patients with health self-assessment 1-2 points

Analysis of the data showed that at the highest level of health self-assessment there is no influence of anxiety level onto the frequency of requests for medical aid. In contrast, at low levels of health self-assessment there was

statistically significant relationship between these parameters. The frequency of medical consultations was statistically significantly different between all 3-anxiety groups. On Fig. 4 there is a graph for patients with health self-assessment 1-2 points.

The same statistically significant ($p < 0,01$) relationship of anxiety level and frequency of medical consultations is observed if patients have three or more somatic diseases (Figure 5). At one registered somatic disease no dependence is detected between the frequency of medical consultations and patients anxiety level.

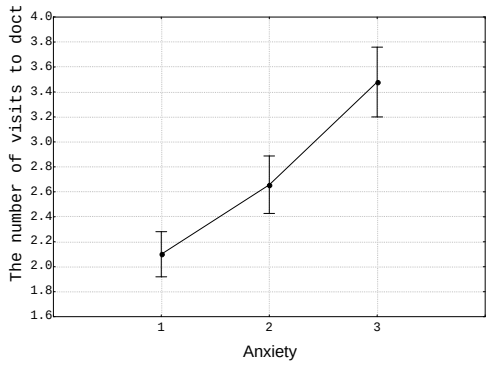


FIGURE 5. Graph of means and confidence interval (95.00%): number of medical consultations at different levels of anxiety in patients with 3 different diseases or more

The relationship between the frequency of medical consultations and the level of depression in patients with different health self-assessment and different number of reported cases was investigated. The analysis showed the following:

- results of men and women with a high level of health self-assessment (3-4 points out of 4) and with no or one disease showed no relationship between the frequency of medical consultations and the level of depression;
- results of women with low health self-assessment (1-2 points out of 4) showed the frequency of medical consultations to be significantly lower with low depression (group 1) than with moderate depression (group 2) and high depression (group 3) (Figure 6a);
- results of women with 3 or more diseases showed the frequency of medical consultations to be differed significantly in all groups with different depression, and it was lower in patients with low depression (group 1) (Figure 7a);
- results of men with low health self-assessment (1-2 points out of 4), the frequency of medical consultations was significantly higher in cases with high depression (group 1) than in cases with moderate (group 2) and low depression (group 3) (Figure 6b);
- results of men with 3 or more diseases showed the frequency of medical consultations to be statistically significantly higher at the high (group 3) rather than at low (group 1) depression (Figure 7b).

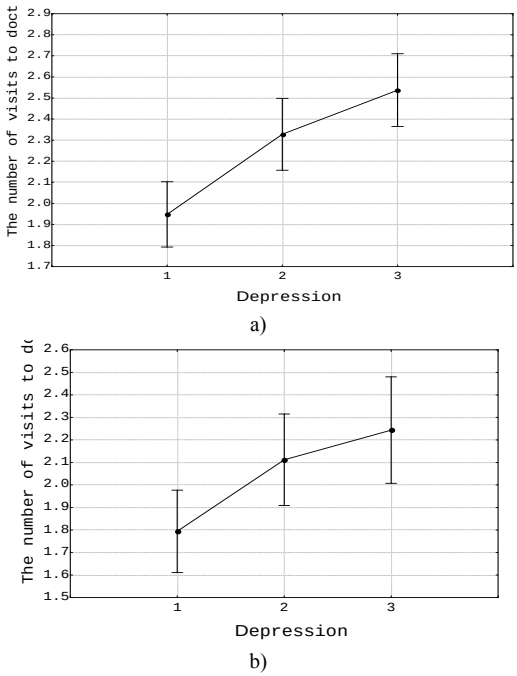


FIGURE 6. Graph of means and confidence interval (95.00%): number of medical consultations at different levels of depression among women (a) and men (b) with low health self-assessment 1-2 points

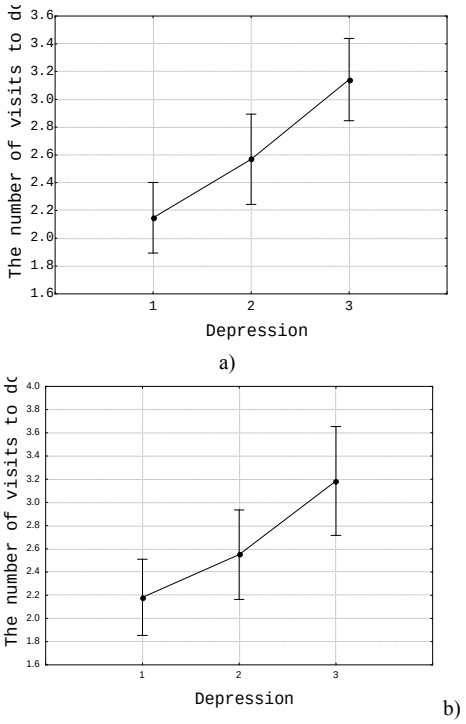


FIGURE 7. Graph of means and confidence interval (95.00%): number of medical consultations at different levels of depression among women (a) and men (b) with 3 and more diseases

These data indicate that the workload of medical staff can be determined not only by the level of disease, but also by psychiatric features of patients.

5. Discussion

The main purpose of medical information systems is ensuring effective information support of all members of the healthcare system in the process of management and of health care delivery.

In the present study there was confirmed the hypothesis that the personal characteristics of patients may have an impact on all phases of healthcare.

Economic Optimization of health depends on the timeliness and accuracy of diagnosis of diseases and their prevention, as well as the cost of patient management. An important prevention strategy is a permanent population health screening. The conduct and improvement of screening methods may not seem beneficial in some cases, but in general, the mass screening tests provide improved population health, if they meet certain principles [18]. However, the effectiveness of screening programs should be scrutinized and validated before implementation [13].

Population screening is not possible without the use of questionnaire methods. However, the study showed that the identification of diseases through self-assessment by questionnaire in certain situations can lead to significant errors and not reveal underlying disease or predict the probability of their occurrence. These issues must be considered when designing the EHR, as comprehensive qualitative data, which allow to develop and use complex methods to solve medical problems for the development of efficient algorithms for the prevention, diagnosis, and treatment are necessary.

Identified potential errors in the diagnosis of disease suggest that to improve the quality of diagnostic health monitoring it is necessary to evaluate personal characteristics of patients and take them into account at all stages of healthcare. Taking into consideration personal characteristics of patients, along with the dynamics of the recorded parameters should increase the sensitivity and specificity of diagnostic procedures, and efficiency of patient management. Computerization of healthcare industry, on the one hand, made it possible on the basis of the accumulated data build intelligent systems for management of patients. However, on the other hand, the construction of intelligent medical systems is impossible without a systematic approach to the assessment of the patient, which is, taking into account not only clinical, but also other personal characteristics of the patients.

Economic benefits of accurate diagnosis, including screening, conducted by MIS, and therefore based treatment will exceed the cost of developing high-quality information systems and a full diagnosis procedure [16]. Since in this case it will be possible to shorten the disability period, reduce the cost of beneficiary drug coverage, claims, various allowances, the disability cases, etc.

MIS, in particular, EHR with proper design and exploitation can be used to generate secondary data, which are necessary for the improvement of all medical procedures of forecasting before the treatment and rehabilitation of patients [19].

Results presented in this paper revealed only some of the problems that are necessary to be considered in the design and use of MIS. In particular, EHR developers should be focused on the expansion of the recorded performance and automation of complex assessments of patients with the use of a wide range of parameters, including personal factors. *It was shown that healthcare staff workload depends on the mental characteristics of the patients.* The data obtained are taken into account in implementation of the Active Longevity portal.

Therefore, for the collection of patients' personal characteristics data tests that determine some personal characteristics of patients were included into the system. This portal is currently used by professionals to collect and prior analysis of medical research data. Specialists who advise physicians in planning examinations recommended to use SF-12 test questionnaire as mandatory to determine anxiety and depression. In the future, we plan to expand the research methods of personality characteristics of patients in order to select the most relevant solutions for specific health problems.

6. Conclusions

The research showed that such indicators as anxiety, Physical Component Score (PCS) of SF-12 test and Kettle's index (weight-height index) can influence onto the probability of error in the identification of effort angina in screening surveys of patients and in the use of health self-assessment questionnaire. At low levels of anxiety, self-esteem does not reveal the effort angina in 83.1% of cases, and at the high level in 43.3% of cases with CVD patients. At a high level of Physical Component Score the error of effort angina detection was in 78.6% of cases, with low PCS in 31.0% of patients with CVD. Besides with the normal Kettle's index in health self-assessment CVD was not revealed in patients with this disease in 83.9% of cases. Based on the results it can be stated that the personal characteristics of patients may have a significant impact on the results of screening and preventive medical examinations, which use rapid surveys of patients and methods of self-assessment questionnaires. The necessity of usage of personal characteristics indicators of patients in clinical practice is also confirmed by other results of the study. Thus, it was revealed the influence of the level of anxiety and depression on the frequency of occurrence of cardiovascular disease in hereditary background of CVD; those patients with higher levels of anxiety and depression are much more likely to develop cardiovascular disease.

The results support the assumption that the effectiveness of MIS in solving medical problems such as the prevention, diagnosis, rehabilitation and forecasting, will depend on the inclusion into the system of both

clinical and personal characteristics of the patients. The results show that *personal characteristics of patients affect cost-effectiveness of healthcare:* with a lower health self-assessment and higher levels of anxiety and depression patients are more likely to seek medical help. Such behavioural characteristics of patients are not taken into account when planning medical staff workload, which can lead to higher actual workload of medical staff and health facilities. The results suggest that the lack of indicators collected in the databases of medical information systems can lead to serious mistakes in planning and management processes of healthcare.

Acknowledgement

The research was done in the framework of the scientific and technical program of the Ministry of Health of the Republic of Kazakhstan (Contract N 168).

References

[1] Millennium Development Goals: progress towards the health-related Millennium Development Goals. Fact sheet, N 290, May 2011 http://www.who.int/mediacentre/factsheets/fs290/en/index.html
[2] Japan Population Pyramid. http://www.nationmaster.com/country/ja/Age_distribution
[3] World population projections. Growing pains. May 5th 2011, The Economist online http://www.economist.com/blogs/dailychart/2011/05/world_population_projections
[4] ICT & Ageing – Final Report, January 2010, ed. by L. Kubitschke and K. Cullen (WRC), empirica Gesellschaft für Kommunikations- und Technologieforschung mbH. Oxford Str. 2, 53111, Bonn Germany
[5] Muhamedyev R. I., Dmitriyev V. G., Maratov M. M., Ualiyeva I. M., Taishmanov B, Muhamedyeva E. L, Zagulova D. V., Mansharipova A. T. (2012) The web portal "Active longevity of Kazakhstan population": actuality, objectives, functions and preliminary results. *The 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems*, SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, 978-1-4673-2743-5/12/$31.00 ©2012 IEEE 571-576
[6] Islamov S. (2010) *Open Access Research Journal, Medical and Health Science Journal,* **1**(1), 59-65
[7] Zigmond A. S., Snaith R. P. (1983) *Acta Psychiatr Scand*, **67**(6), 361-70.
[8] Rose G., McCartney P., Reid D. D. (1977) *Br. J. Prev. Soc. Med.*, **31**(1), 42–48
[9] Roca J., Alonso A., Hernandez C. (2011) Integrated Care for Management of COPD Patients: Time for Extensive Deployment, *Meditsinskie tehnologii*, The Russian State Medical University (RSMU), **4,** 76-81
[10] Higgins J., Salbach N. M., Soicher J., Jaglal S. (2004) *Journal of the American Medical Informatics Association,* **11**(6), 514–22
[11] Mayo N. E., Nadeau L, Levesque L, et al. 2005 *Med Care*, **43,** 1194–202
[12] Rich M. W., Beckham V. et al. (1995) *N Engl J Med,* **333,** 1190-5
[13] World health statistics 2012, World Health Organization 2012, http://www.who.int/gho/publications/world_health_statistics/EN_WHS2012_Full.pdf, pp. 35.
[14] Oortwijn W., Banta H. D., Cranovsky R. (2001) *Int. J. Technol. Assess Health Care,* **17**(3), 269-74
[15] Albert C. M., Chae C. U., Reprode K. M. (2005) *Circulation,* **111,** 480-7
[16] Lohanova Z. M., Vasilenko V. V., Timofeev C. E. (2010) Criteria for the diagnosis and staging of hypertension angina pectoris: theory and practice. *Russian Medical Journal*, Publishing House "RMJ", **18**(22), 1393-6
[17] Wilhelmson K., Rubenowitz L. E., Andersson C., Sundh V., Waern M. (2006) *Aging Clin Exp Res.,* **18**(1), 25-33
[18] Wilson J. M. G., Jungner G. (1968) *Principles and practice of screening for disease* WHO Chronicle Geneva: World Health Organization. Public Health Papers. 34 p.
[19] Madabhushi A., Agner S., Basavanhally A., Doyle S., Lee G. (2011) *Comput Med Imaging Graph,* **35**(7-8), 506-14

Authors

Diana Zagulova, born 20.09.1964, Riga, Latvia

Current position, grades: dr. med., prof.
University studies:
1 - The Baltic International Academy, Riga, Latvia
2 - Tomsk State University, Tomsk, Russia
Scientific interest: information systems, systems theory, comprehensive health diagnosis
Publications: 63
Experience: about 25 years

Ravil I. Muhamedyev, born 23.06.1959, Irkutsk, Russia

Current position, grades: professor, head of department
University studies: International IT University, Almaty, Kazakhstan
Scientific interest: machine learning, simulation of anisochronous systems
Publications: about 130

- Integration of Results of Recognition Algorithms at the Uranium Deposits. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.18 No.3, 2013
- Machine-learning techniques in pattern recognition rocks at uranium deposits. / Proceedings of the National Academy of Sciences of Kazakhstan, 2013, №3. С.82-88
- Geophysical Research of Boreholes: Artificial Neural Networks Data Analysis. The 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, 978-1-4673-2743-5/12/$31.00 ©2012 IEEE, p.825-829

Experience: about 25 years

Alma Mansharipova, born 13.09.1962, Almaty, Kazakhstan

Current position, grades: Director of the Department of Science
University studies: Kazakh Russian medical university
Scientific interest: research in the field of medicine, cell biology, bioinformatics
Publications:

- Therapeutic regression of atherosclerosis Atherosclerosis (supl). 2011, **12**(1), 70
- Prognosis of non Q-wave myocardial infarction in the elderly residing in Almaty Medical and Health Science journal. 2012. **10,** 10
- Variant Polymorphisms of GST and XRCC genes and the early risk of age associated disease in Kazakhstan World academy of science, engineering and technology. Internal Journal of medical, pharmaceutical science and engineering. 2012, **6**(11), 15-20

Experience: 24 years

Irina Ualiyeva, born 21.08.1969, Republic of Kazakhstan, Kostanay

Current position, grades: Assistant Professor, International IT University, Almaty, Kazakhstan
University studies: International IT University
Scientific interest: data mining, machine learning
Publications:

The web portal "Active longevity of Kazakhstan population": actuality, objectives, functions and preliminary results. The 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, 978-1-4673-2743-5/12/$31.00 ©2012 IEEE, pp. 571-576

Experience: about 20 years

Elena Muhamedyeva, born 19.05.1971, Riga, Latvia

Current position, grades: PhD-student of Riga Technical University (Riga, Latvia), researcher of International IT University (Almaty, Kazakhstan)
University studies: Riga Technical University, International IT University
Scientific interest: software engineering, data mining
Publications:

- The web portal "Active longevity of Kazakhstan population": actuality, objectives, functions and preliminary results. The 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, 978-1-4673-2743-5/12/$31.00 ©2012 IEEE, pp. 571-576
- Geophysical Research of Boreholes: Artificial Neural Networks Data Analysis. The 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012, 978-1-4673-2743-5/12/$31.00 ©2012 IEEE, p.825-829

Experience: About 5 years

IT investments justification based on the business driver tree

M. Fay, V. Grekul*, N. Korovkina

National Research University Higher School of Economics

20 Myasnitskaya Ulitsa, Moscow, 101000, Russia

Received 25 December 2013, www.tsi.lv

Abstract

Choosing the right IT project for supporting the company business development is nowadays one of the most critical tasks in information technology management. No one has yet managed to create one optimal solution, equally suitable for different types of stakeholders (business owners, managers, investors). This article, based on the ideas of Value Based Management and business/value-driver trees, concentrates on an original approach to managing investments in enterprise architecture IT component. The method suggested has been successfully applied to evaluate the IT project portfolio within a large metals company in Russia.

Keywords: Value Based Management, IT-project, IT investments, business driver, IT projects portfolio, investments justification

1. Introduction

In 2012 Harvard Business Review published the results of a survey [5] conducted among the top-managers of the world leading companies. According to HBR, 72% of respondents stated that "the use of technology has helped to create new value in business areas" – and also noted that IT departments have become full-scale players able to influence company's competitive advantages and profits. IT projects go after investments playing on the same field not only with other IT projects, but also with the core operations projects. Due to this fact, the objective of justifying the planned IT project outcome (and, needless to say, choosing it) has becoming more and more relevant, especially in connection with the business outcomes and business goals.

According to Gartner Group [3], 35% of all IT expenditures would be managed outside of IT departments. Independent IT projects would be budgeted by business specialists – which It means their strategic importance would be the key focus. However, economical methods of IT investments justification, used by the majority of enterprises, are mostly based on the forecasts and are not able to preview all the possible risks and take into account company strategy and tactics.

One of the latest researches conducted by the Russian ITSMF-forum [6] note that entrepreneurs tend to invest "spontaneously" and 3 out of 4 CIOs report on the non-comprehensive current justification methods that will have its negative impact in a long-term period. "Business-oriented IT assets" is a goal stated by many companies but, unfortunately, in most cases this will not come true.

Multinational experience and best practices can help in such a situation to define and consider companies' business priorities in terms of "business drivers". Without doubt, IT investment planning should be systematical and based on the contribution to the business-oriented information services provided to the company. Let us define the "information service" here as "a way of presenting the IT value to the users by helping them in achieving the desired results".

This term appeared back in 2007 in ITIL 3rd edition and only now it is gaining its well-deserved popularity among Russian CIOs, while it is one of the most direct ways to link together the company value growth factors (business-drivers), IT services offered to the business and IT systems/infrastructure costs. For instance, the project of creating a corporate web site can (and should!) be regarded NOT as a simple web-site creation (which is a method), but as creating "a company office/ representation" online (which is the primary objective). Thus, the business service provided can be seen from the point of view of improving the company image and brand recognition (they could already be examples of business-drivers).

According to the researches of the practices of evaluation and IT benefits management for SMEs, provided by Edith Cowan University and Brunel University professors [2], the lack of a holistic model for IT investment management is likely to result in several consequences:

- ✓ Refusing the IT infrastructure investments (with negative outcomes in the long-term period);
- ✓ "Intuitive" and "spontaneous" investments;
- ✓ Not taking into account the business and industry specifics – and thus, losing potential competitive advantage.

* Corresponding author: grekoul@hse.ru

Many monographs published at some point highlight these questions. From the scientific point of view really important is the research done by Love Peter E.D. and Irani Zahir [7]. They take a close look at the key difficulties that companies have while justifying the projects, the practices they use to define the strategic/tactical/operational IT outcomes as well as the most popular reasons for positive decisions on IT system implementation.

The problems formulated above demand the creation of IT investment justification method that would rely on both the strategy and business-drivers for defining the IT projects business value. It is the basis for the current research relevance. Important to mention, this article does not pose the aim of describing the existing or creating new and universal ways of defining the strategic goals and IT assets necessary. The article concentrates on the common both for business and IT area of business drivers, proposing a mechanism for their implementation to IT investment justification process.

The key objective of the current research is ensuring the business priorities consideration for the IT projects portfolio creation process through the development of a complex IT investments justification method for the large Russian metals company, whose headquarters is based in Russia with several foreign branches.

2. IT Project Justification Method Components

As already mentioned above, Gartner Group research shows that to the year of 2015 corporate IT expenditures would continue to become the area of expertise and decision-making for business specialists. To define the criteria for the suggested approach adequacy we shall turn to the results obtained by a group of Russian researchers and practitioners, united in the *"IT value"* project aiming at developing the topic of information systems economic efficiency. Their findings define the key problems mentioned by managers on the topic of IT [1] (Table 1).

TABLE 1. Key problems in evaluating IT investments

Business stakeholders	Highlighted problems
Entrepreneurs	Low comprehension and consideration level of business and industrial specifics
Investors	High rate of errors for cash flow forecasts for IT.
Top managers	Lack of IT evaluation possibilities without the deep understanding of business strategy
Line managers	Lack of investment business effect possibilities and inability to formulate the requirements for the investment projects

Analysis of these results enabled us to define the following criteria for the method to be developed:

1) Relying on the best practices that take into account company strategy;

2) Having the opportunities to define the potential benefits from IT projects in business terms;

3) Assembling the IT projects portfolio based on the strategic priorities.

2.1. RELIANCE ON THE BEST PRACTICES THAT TAKE INTO ACCOUNT COMPANY STRATEGY

In the VBM concept, graphical model of which is presented at Figure 1, the drivers are considered to be 'the performance variables', influencing 'customer satisfaction, cost, capital expenditures, etc.

However, as suggests the manual published by McKinsey, who originally introduced the method in their quarterly review, generic value drivers might apply to most business units. But the lack of specificity forced us to make adjustments to the model by adding the non-financial factors as well.

Due to the results of the annual report analysis (especially the part concerning company's strategic objectives), six interviews, conducted with business and IT managers, and the consideration of the IT annual budget (based on the project portfolio) the information gathered was sufficient to build a business-driver tree.

To begin with, it is important to highlight ABC's strategic priorities. Company's branches are located in different countries and continents with the focus on the USA's market while during this project realization the enterprise was preparing to conduct the IPO. Top managers named several critical tasks, such as:

- ✓ Standardization of the key business processes at the level of business divisions;
- ✓ Establishing effective cooperation between production and management units as well as between geographically distant branches;
- ✓ High business flexibility to quickly adapt to the changing market conditions.

Company's economic value optimization is supported by several drivers, providing detailed description of the Level-1 VBM model:

- ✓ Revenue: Market coverage and Market share;
- ✓ Costs: Production and SG&A costs;
- ✓ Working capital: Operational investments and Macroeconomic factors;
- ✓ Fixed capital: Net assets value and Assets turnover.

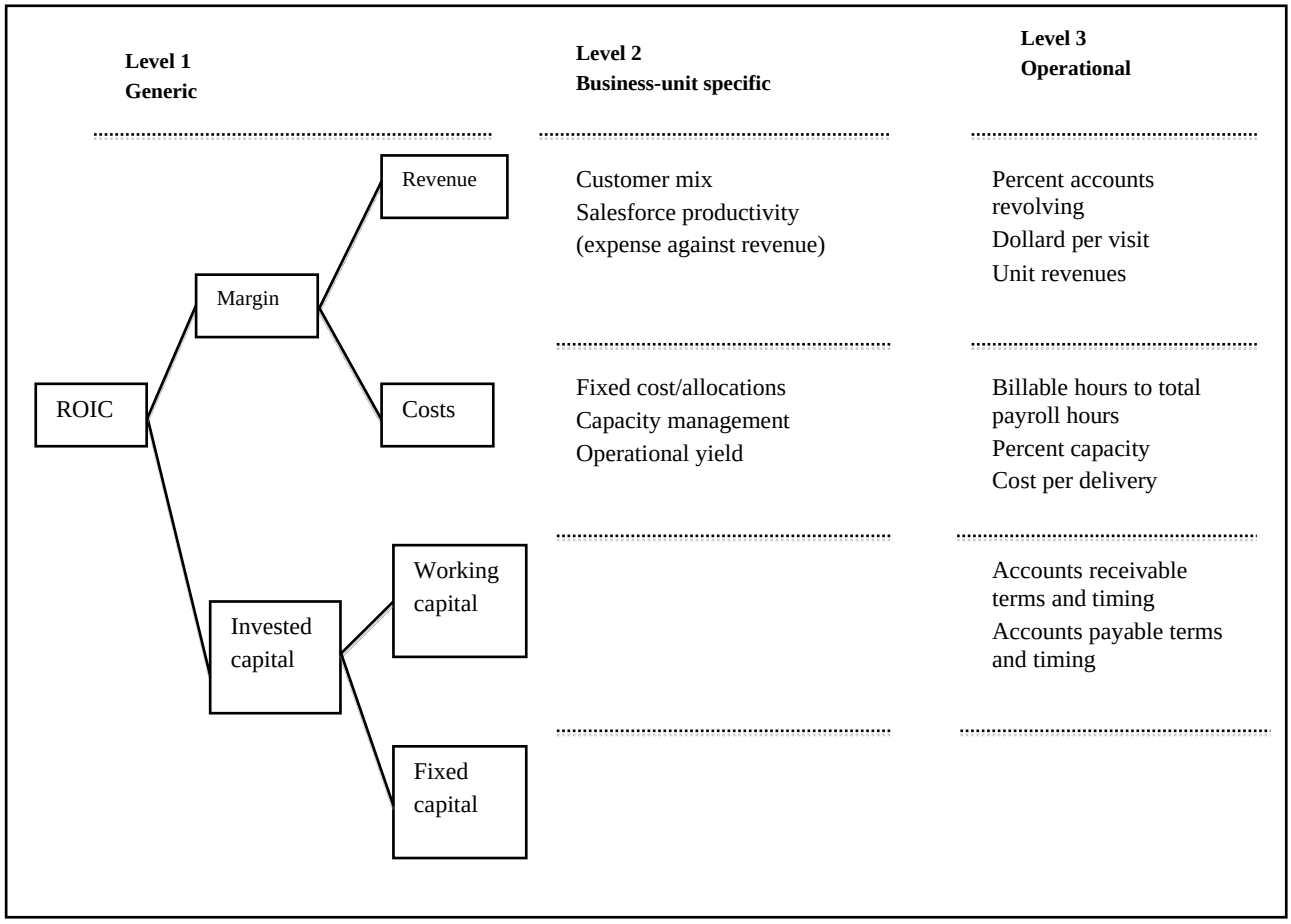


FIGURE 1. VBM value driver tree model

Company's strategic value optimization is supported by several drivers, extending the original VBM model:

- ✓ Proactivity;
- ✓ External optimization;
- ✓ Focus on perspectives;
- ✓ Analytics efficiency.

Based on whole business modelling and analysis activities conducted the driver generic extended tree was built (Figure 2), so let us now turn to its potential application in the IT justification process.

2.2. FORMULATION OF THE POTENTIAL BENEFITS FROM IT PROJECTS IN BUSINESS TERMS

Another problem, associated with evaluation of program realization gross profit, is refusing the projects that do not reach the critical values in terms of NPV and ROI. However, if the company management estimates the highlighted business services as critical, this difference between the actual and target values can be remitted. For instance, if the costs for the urgent organization of telepresence facilities for some important events have not been planned and justified, the project would be rejected without perhaps even being considered. At the same time, with the business-drivers tree expansion proposed the managers just need to justify that the non-financial benefits provided by telepresence outweigh the need for reaching the critical financial values. In the case above, business driver "internal optimization" that could present the objective of improving corporate communications, was considered critical according to the short-term (2-3 years) company goals.

2.3. CREATION OF IT PROJECT PORTFOLIO BASED ON THE STRATEGIC PRIORITIES

IT project portfolio could be formed being based on one more method: ValIT, introducing the whole concept of consideration financial and non-financial benefits as well as risk awareness [4].

Expanding it with the forth factor (outweighing of the financial indicators critical values by non-financial benefits), that replaces the original 'Strategic alignment' factor in Val IT case proposed, let us turn to the matrix formed and applied to practice in the company described above (see Table 2).

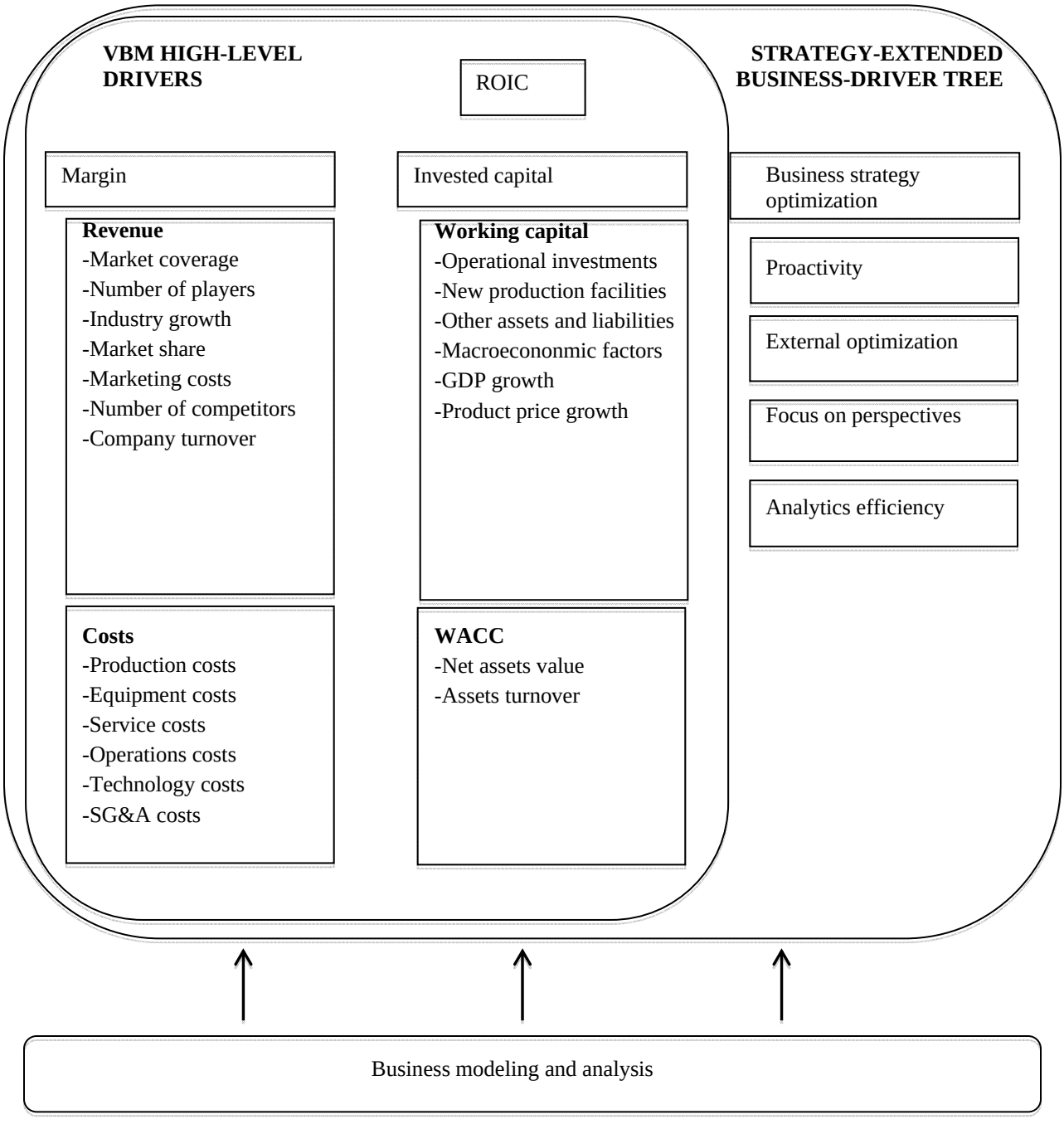


FIGURE 2. Extended driver tree

TABLE 2. IT project consideration matrix

Project	IT project consideration results			
	Calculated risk acceptable?	*Financial targets met?*	*Do non-financial benefits outweigh the need for financial indicators critical values?*	*Non-financial benefits explicit*
Shared service centre	Yes	Yes	-	Yes
RFID implementation to increase the accuracy of data processing and production control	Yes	No	Yes	No
Telepresence facilities to conduct negotiations and meetings with company branches and foreign offices	Yes	No	Yes	Yes

3. Conclusions

The practical results for the work conducted due to the specifics of the project and the topic itself could be evaluated for now only at the qualitative (not quantitative) level and be presented by the customer's reviews. According to the COO of the company where the pilot project implementation took place, *'…earlier the key technological reason for IT projects was the necessity to change the outdated legacy systems and make an upgrade, while now the primary goal is maintaining the alignment with current business priorities preserving flexibility of a system landscape transformation for the business requirements.'*

As the method proposed has been already adopted by several other enterprises, the first steps towards the professional business-oriented view on IT have definitely been taken and the full-scale cooperation between business and IT would not be long in coming.

References

[1] Anan'yin V., Zimin K., (2012) *IT value, Theory and Practice* CIO Congress Kirov http://bit.ly/139QkWJ 28 Apr. 2012
[2] Alshawi S, Irani Z, Baldwin L 2003 *Benchmarking*, **10**(4), 414-423
[3] *Gartner Reveals Top Predictions for IT Organizations and Users for 2012 and Beyond* http://www.gartner.com/newsroom/id/1862714 1 Dec. 2011.
[4] *Enterprise Value: Governance of IT investments. The Business Case. IT Governance Institute* 2006 http://www.isaca.org/Knowledge-Center/Val-IT-IT-Value-Delivery-/Documents/VAL-IT-business-case.pdf 20 Oct. 2013
[5] *Innovation from Efficiency: Unlocking the Value in IT Systems. Harvard Business Review Analytic* http://static.hbr.org/hbrg-main/resources/pdfs/tools/17441_HBR_Dell_Report_webview.pdf 1 Feb 2012
[6] *ITSMF-forum conference*, http://www.itsmfcon.ru/2013/ 12 Sep. 2013
[7] Love P. E. D., Irani Z 2004 *Information & Management,* **42,** 227–242

Author

Maria Fay, born 17.06.1989, Kirov, Russia

Current position, grades: HPI School of Design Thinking (Potsdam, Germany)
University studies: Master of Business Informatics, National research university – Higher school of economics (Moscow, Russia)
Scientific interest: IT investments, IT governance, Technology commercialization
Publications: 3
Experience: Maria has started her professional career in 2010 at Accenture Technology consulting practice, where as a part of a team she developed the enterprise architecture and IT governance solutions for major Russian utilities and production companies. Afterwards she continued her professional path in Skolkovo Foundation strategy department, getting a closer view on the business aspect of IT investment activities and taking part in CRM system implementation from the customer's side. Winner of organized by ITSM forum all-Russian research contest among graduates of 2013

Vladimir Grekul, born 15.09.1949, Sakhalin, Russia

Current position, grades: professor, National Research University Higher School of Economics, Moscow, Ph.D.
University studies: Kiev Higher Engineer Aviation Training School of the Air Force
Scientific interest: IT management
Publications: 90
Experience: V. Grekul has 40 years experience in design and implementation of information systems. He has been working at NRU HSE since 2001 as Professor, Head of the Department of Corporate Information Systems

Nina Korovkina, born 24.10.1947, Moscow, Russia

Current position, grades: Associate professor for the Corporate information systems department, National research university – Higher school of economics (Moscow, Russia)
University studies: Moscow institute of railway transport, engineer-mathematician specialization majoring in "Mathematical and computing devices"
Scientific interest: Design and implementation of information systems, IT project management
Publications: 34
Experience: During her professional career, Nina Korovkina worked in the Central economics institute for economics and mathematics (division of the Academy of sciences, Moscow, USSR) as well as in the Main computer centre of State planning committee. Since 1994 she has been working in the National research university - Higher school of economics and took part in the projects carried out for the Ministry of Education, Ministry of Economic Development and Trade, Ministry of Information Technologies and Communications.

CBM-based integrated management information system design for mine construction enterprises

Bang-Jun Wang *

School of Management, China University of Mining & Technology, Xuzhou of Jiangsu, 221116, China

Received 1 January, 2014, www.tsi.lv

Abstract

To solve existing problems in the management for mine construction enterprises, this research introduced the concept of CBM (Comprehensive Budget Management) on the basis of data and information demand analysis using enterprise management decision. Next, under the circumstance of project management, the market mechanism was introduced in mine construction enterprises, materializing the CBM-based integrated management information system design for mine construction enterprises. The system architecture encompassed six modules, namely, production progress management, project material management, mechanical and electrical equipment management, human resource management, integrated cost management and performance management. The system covered the integrated management information system for each process of the mine construction management. With the B/S structure, technological development approaches of this system consisted of UML modelling technique, dynamic configured technology, database design and implementation. The actual application of integrated management system in sample mine enterprises showed optimized enterprise management process and improved data processing proficiency, greatly enhancing the financial performance and competitiveness of mine enterprises.

Keywords: Comprehensive Budget Management (CBM); mine construction enterprises; integrated management information system; design and application

1. Introduction

CBM (Comprehensive Budget Management) enables enterprises to optimize the resource allocation, analyse, coordinate and control the budget implementation. Specifically, CBM facilitates the analysis, prediction and decision-making about sales and profits, production, costs, expenses and funds focusing on strategic targets of enterprises. The aim of CBM is to systematically and efficiently coordinate and implement business activities of enterprises [1, 2]. Chandler (1977) considered that CBM plays a crucial role in the maturing and development of modern business enterprises in the West [3]. CBM is widely recognized as a management system that greatly drives maturing and development of modern business enterprises, and a main method of internal management and control for enterprises. As stated by David Otley, CBM is one of the few management methods that can integrate all key problems in an organization into one system. Since the invention and application of CBM in enterprises like GE, Dupont and GM in America in 1920s, it has soon become a standard operation procedure of large business enterprises [4]. From the initial functions in planning and coordinating, CBM is now an integrated management tool that combines enterprise management strategies, enterprise control, motivation and evaluation [5, 6, 7]. Currently, CBM has become the core of management strategy of enterprises [8, 9].

With the gradual completing of China's market economy system, a fair market competition mechanism has been progressively established. Nowadays, mine construction enterprises are experiencing intensifying fair market competition mechanism [10]. The conventional business system based on administrative planning constrains the rapid development of mine construction enterprises to a large extent, and the restructuring of the business management structure of mine construction enterprises becomes inevitable [11]. The implementation of CBM based on project management becomes an irresistible trend of internal management for mine construction enterprises. Accelerating the restructuring of the management system for mine construction enterprises could facilitate the realization of competitive advantages in market and overall economic efficiency covering aspects like market scales, costs, prices and quality [12, 13]. From a general perspective, theoretical and application research on CBM among Chinese enterprises present the emphasis on planning and financial management. The systematical study integrated with business is limited [14]. The implementation and budgeting management demonstrates outcomes like insufficient science, seriousness and authority. The assessment outcomes could not be identified as relevant to the employees' benefits and compensation, and it is incomplete to establish feasible and effective incentive and restraint mechanism. Moreover, the budget

* Corresponding author. Tel: +00-086-13813465302;
E-mail: wbjun@cumt.edu.cn

management fails to motivate the initiatives of person in charge effectively, and various problems occur in the process of budget implementation like loosening budget constraints, formality in implementation and difficulty to realize expected management outcomes. The supportive role of the management information system on decisions of enterprise managers need to be further enhanced. Thus, based on the features of mine construction enterprises, the current extensive production mode can be reconsidered and changed because of its conflicts with the requirements of scientific management development for modern enterprises. Thus, it is of great significance to conduct scientific and systematic research on the production process, to integrate management information and to reduce production costs to the greatest extent based on information technology and the latest information management platform.

To this end, on the basis of analysis and summary of current CBM studies, this research introduced CBM in the mine construction enterprises, and integrated researchers' suggestions and experiences about improving internal management to combine CBM with distinct features of mine construction enterprises in China. A set of informatisation platform with project as the structural unit was also designed for mine construction enterprises. This research could provide effective guidance for mine construction enterprises to comprehensively and systematically implement CBM.

2. Features of the System

2.1. ANALYSIS AND RESTRUCTURING OF INFORMATIZATION-DRIVEN PROCESS

The principle of the optimization design for mine construction enterprises was to realize the restructuring of enterprise businesses under the collaboration of market chains and to materialize the optimal human-project combination and product/service performance. In this way, each process could directly serve for the project object. Besides, each basic decision unit was assigned with corresponding subsidiary autonomy, and the outcome of each process could be measured by currencies. This could further evaluate the quality, costs and benefits, and optimize the internal resource allocation via the structural integration, business process integration and resource integration. The optimization design facilitated the reduction of construction costs and the improvement of beneficial efficiency in enterprises.

2.2. SCIENTIFIC SYSTEM DESIGN AND COMPREHENSIVE FUNCTION INTEGRATION

This integrated management information system combined powerful remote applications and realized synchronous transmission for cross-regional management resource data in each project department, which would be analysed collectively in the central data bank. The advanced frame-based software technology ensured the flexibility, reliability and extendibility in the research and development. Distributed data processing of SQL Server and Internet/Intranet also promoted the centralized enterprise resource management. The all-informatisation management performed excellently in resource sharing, smooth information flow and intelligentised internal logistics, and materialized remote management and mobile support.

2.3. FLEXIBLE MODULE DESIGN AND CUSTOMIZED SERVICES

All functions of this system were designed on the basis of modules so that users could flexibly configure each module to their needs. The efficient data processing ability and complete data backup mechanism effectively reduced the network flow during data transmission, which shortened the system response time and guaranteed the effectiveness and accuracy of data. Users could easily operate the system after training.

2.4. ADVANCED TECHNICAL DEVELOPMENT SYSTEM AND GUARANTEED SYSTEM QUALITY

In the research and development of this system, strict process control was introduced for all internal units, and object-oriented UML modelling technique was applied in the design. In addition to the introduction of user-oriented design and UML-based depiction of all design models, RUP (Rational Unified Process) was introduced in the internal management for mine construction enterprises, which could guarantee the development quality and maintainability of the software system.

3. System Structure Design

This management information system improved the internal marketization for mine construction enterprises, their budget management, compensation management and implementation details of evaluation, and promoted CBM for mine construction enterprises. A range of modules were provided by the system, including production progress management, project material management, mechanical and electrical equipment management, human resource management budget management, and integrated management statements etc., which is shown in Figure 1.

Internet was used as a support for the system to ensure the information exchange within enterprises. The system network architecture and hierarchical structure is shown in Figure 2.

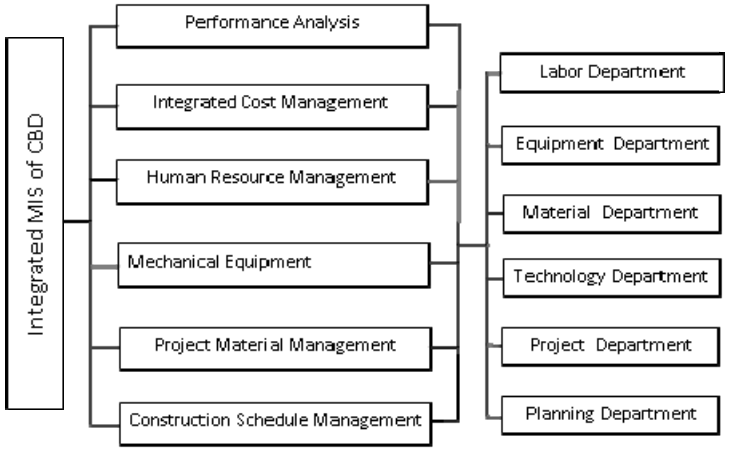


FIGURE 1. System structure diagram

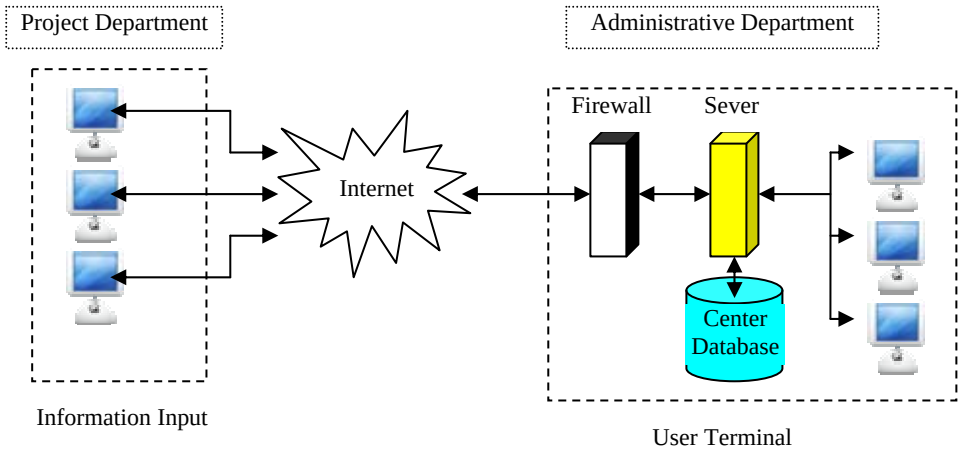


FIGURE 2. System network structure

4. Main Modules and Functions of the System

4.1. CONSTRUCTION SCHEDULE MANAGEMENT

The production progress management for the Project Department encompassed the following parts: daily construction report management, original records of construction units and scheduling information query. This module automatically produced daily reports, weekly reports and monthly reports of construction schedules. By the comparative analysis with the plans, this module dynamically controlled the implementation scheduling. Functions of this part were as follows.

4.2. PROJECT MATERIAL MANAGEMENT

Project material management covered material information, warehouse location, inventory / loss reports, material allocation / storage, cargo-dropping management and material information query. This module automatically formed dynamic statements of material statistics for the Project Department.

Functions of this module included stock material input, inventory and cargo-dropping, material inventory query, maternal balance query, monthly material statements and so on. The material plan management covered annual plans, monthly plans, supplementary plans, temporary plans and emergency plans. The material staff of each Project Department declared the above mentioned plans of material requirement and input the information in the system. The plans would be assessed by related departments. The system then collected all approved plans and formed the general plan of all material needs for construction enterprises.

The inventory input and cargo-dropping mainly registered the detailed stocking and cargo-dropping information. The data was to take inventory of materials. The material inventory and balance conditions referred to the statistics of all materials in the warehouse. The function gradually collected material statements of the Production Project Department and complete monthly material statements for the enterprises. The statements would be output according to the requirements to form the material costs in the expenditure part.

4.3. MECHANICAL AND ELECTRICAL EQUIPMENT MANAGEMENT

Based on the inventory analysis of mechanical and electrical devices, equipment coding was conducted to produce a variety of statements for integrated query of mechanical and electrical devices, basic information of devices, inventory and cargo-dropping management, device repair and acceptance, device allocation and management, usage monitoring, equipment scrapping management and device disbursement and settlement. The module functioned in effective and comprehensive management of all static and dynamic information in the full life circle of devices. Device status and distribution at different work sites were dynamically tracked, which presented accurate records, tracking and controlling of device-related processes, such as device purchasing, rental, examining, repairing, maintaining, and allocation of spare parts. Scientific and orderly management of devices could fully meet the production requirements, and provide proofs for direct expenses on mechanical devices in the cost management.

4.4. HUMAN RESOURCE MANAGEMENT

The module of human resource management is mainly provided freely to set enterprise's organizational structure. Under CBM managerial frame, information of functional departments and the project department such as basic information of employees, academic background, resumes, archives, contracts and other information were systematically integrated. This module also provided simple and direct logging in, query and modification. The employee attendance management recorded the attendance of each work shift of each construction teams and groups, providing a basis for labour costs in performance evaluation.

The module was founded on human resource planning and management, covering aspects like institutional framework, positions, human resource management, labour contract, recruitment, benefits, training and development, attendance and performance appraisal and remuneration. Meanwhile, this system could promote the automation and optimization of HR management process in an all-around way. Its primary functions included employee information management, compensation management, performance appraisal and management, labour management etc.

4.5. INTEGRATED COST MANAGEMENT

Integrated cost management module realized plan management of costs on material, labour, mechanical devices, and other items, and analysed the implementation of budget plans. The budget management sub-system realized budget management at all levels. With the logging and maintaining of cost items, a budget data platform within enterprises including specialized market items, specialized items, project budget of construction groups / teams. This module conducted data collection and analysis of device rental conditions. According to expenses of using construction machinery in projects, the overspending and surplus conditions of machinery costs could be gained. The module also developed and tracked the implementation of material requirement plans, thus dynamically realizing the rational purchase/inventory and distribution controls of materials. Moreover, this module achieved real-time tracking and assessing material consumption at each basic unit (project departments and construction groups), integrated analyses of costs on materials and devices, and accomplished overall and individual costs of at each basic unit and timely uploaded the data.

4.6. PERFORMANCE ANALYSIS

This module integrated information of the above modules, formed a complete business summary sheet for mine construction enterprises, and conducted business analysis. The business analysis was targeted at enterprise business states of each project department and summarized business states according to years and months. Functional departments summarized and analysed specialized costs, overspending and surplus conditions of comprehensive costs and the specific reasons according to the uploaded information of costs. Lastly, it would form performance analysis statements.

(1) the Performance of Project Department

$$\mathrm{TR}_{pd} = \sum AUP * UBP . \tag{1}$$

In formula (1), TR_{pd} is total revenue, AUP is accepted unit projects and UBP is unit budget price of mine construction enterprise;

$$\mathrm{TC}_{pd} = (LC + MC + MUC) * ACODC * ACIC , \\ * ACLIE * ACT . \tag{2}$$

In formula (2), TC_{pd} is total costs of mine construction enterprise. LC, MC and MUC are the abbreviations of labour, material and mechanical usage costs. $ACODC$, $ACIC$, $ACLIE$ and ACT are adjustment coefficients of other direct costs, indirect costs, labour insurance expenses and tax.

$$NP_{pd} = TR_{pd} - TC_{pd} . \tag{3}$$

NP_{pd} is the net profits of each project department of mine construction enterprise.

(2) formula of Mine Construction Enterprise's Operational Performance

$$STR = \sum TR_{pd} + OI \,. \quad (4)$$

Where *STR* is total revenue of mine construction enterprise and OI is other income.

$$STC = \sum TC_{pd} + OCAI \,. \quad (5)$$

In formula (5), *STC* is total costs of mine construction enterprise and *OCTI* symbols other cost adjustment items.

$$NP = STR - STC \,. \quad (6)$$

So, *NP* is net profits, which is key indicator of operational performance of mine construction enterprises.

Functions of this module include: (i)developing involved functions in function application requirements based on all technical, quality control, performance and other plan management indexes, designing cost plan sheets, plan management of cost items of individual item and unit project, forming settlement prices of projects and working procedures; (ii)when the unit price of materials and salaries was changed, the system could automatically recalculate the settlement price; (iii) it could also realize the comparative analysis with actual expense indexes, and analyse the quotas for overspending (surplus) of each item. With the produced variance report, enterprises could carry out corresponding management measures based on the data, thus greatly improving the precision and pertinence of CBM.

5. System implementation and application

5.1. TECHNICAL IMPLEMENTATION APPROACHES

According to the above systematic analysis and design concept, the system adopted the B/S structure, with Windows Server2008, Web server and IIS7.0 as its servers. SQL Server2008 was used as the back-end database. The operation system at the client end was Windows XP. For front-end development tools, VB.Net and C# were selected. With the networking protocol of TCP/IP, the system was designed to facilitate the connection with Internet. Key technologies included UML modelling technique, dynamic configured technology, and database design and implementation technology.

The system was an Internet-based system with Java Applet embedded. Applications were designed at the server end and the client's end only presented necessary system information, which guaranteed the security of procedure codes. The data confidentiality was protected by four levels of security control approaches from operation system, database and application to data operation. Besides, a secondary password technology operated by different roles and different managers was also applied, which enhanced the safety and reliability of data.

5.2. SYSTEM APPLICATION

CBM system has been used in 12 projects of mine construction enterprises such as No.4 Engineering Department of China Coal First Construction Company limited for over 2 years, and its performance was unanimously highly regarded by managers of these enterprises. The system greatly changed the extensive construction management mode of mine construction enterprises into a process starting from project analysis to budget decomposition, then to real-time dynamic input of integrated costs and feedback of business performance. This process not only better fit CBM-based construction mode but also improved the construction efficiency. The mine construction enterprises have witnessed great improvements in aspects like management performance, management mode, management system and methods, management mechanism, management foundation, business process, organization structure, rules and regulations, basic data, information integration and processing, employees' overall competence, decision-making quality, enterprise image and competition etc. At the mean time, the system boosted the innovation of project management.

6. Conclusions

In order to obtain competitive edges in the fierce market comp, mine construction enterprises need to maximize their values, save costs and keep precise profit and loss accounting. With the core of project costs, mine construction enterprises should develop their overall strategy, business strategy and business strategy and strategy, forming strategic budget management and establishing CBM. On the basis of project analysis, CBM adopts cost budgeting as the starting point, cost control as the principal axis of budget control, and costs as the principle indicator of performance evaluation. CBM could facilitate mine construction enterprises to build a new enterprise management system that fits actual conditions of enterprises, materializes orderly management, smooth operation and appropriate monitoring. Besides, it could clearly define responsibilities and rights and coordinates the internal production and business activities by the relationship of market transaction. With CBM, enterprises could minimize diverse forms of wastes and improve their competitiveness significantly.

Acknowledgments

Thank for Mr. V. Zvonarevs and the anonymous reviews' hard work. The research findings of this paper are supported by the Fundamental Research Founds for the Central Universities (2013XK01).

References

[1] Song Lu-Ping (2011) Implementation of a comprehensive budget management to promote the enterprise's strategic target. *China Management Informationisation*, **3,** 50-1 (in Chinese)

[2] Zhang Jian-Ping, Wang Na-Na (2011) Research on comprehensive budget management of enterprise group. *Social Sciences Journal of Universities in ShanXi* , **1,** 35-7 (in Chinese)

[3] David Otley (2010) Research Assessment in the UK: An Overview of 1992–2008. *Australian Accounting Review*, **20**(1), 3-13

[4] Chandler A. D. (1977) *The Visible Hand: The Managerial Revolution in American Business.* Harvard University Press

[5] Chen Chang-Chun, Niu Ke-Hong, Wang Qi-Xiang (2006) *The comprehensive budget management planning and Practice.* Peking: China Economic Publishing House (in Chinese)

[6] Dai Xiao-Hong (2008) On the difficulties analysis, implementation and the solution of comprehensive budget management in the Construction enterprises. *Building Construction*, **3,** 234-6 (in Chinese)

[7] He Li-Rong (2009) On the comprehensive budget management of construction enterprises. *Finance Accounting For Communications*, **4,** 46-8 (in Chinese)

[8] Li En-Zi (2009) Thoughts on the Comprehensive Budget Management Introduced to Geological Exploration Units [J]. *Natural Resource Economics of China*, **9,** 39-42 (in Chinese)

[9] Psuty N. P., Spahn A., Silveira T. M. (2013) Sediment budget as a driver for sediment management at plumb beach, New York USA: vectors of change and impacts. *Journal of Coastal Conservation*, online Springer. DOI 10.1007/s11852-013-0293-9

[10] Yang Jin-Bo (2009) The comprehensive budget management of construction enterprises Based on the discussion of the project responsibility cost. *Money China*, **1,** 166 (in Chinese)

[11] Tang Xiao-Fang, Tao Ping (2012) The enterprise group overall budget management problems and Countermeasures. *Storage Transportation & Preservation of Commodities*, **3,** 168-9 (in Chinese)

[12] Tang Ping (2012) The preparation of China's construction enterprise comprehensive budget. *China Business Update*, **4,** 223-4 (in Chinese)

[13] Cao Na (2012) The comprehensive budget management in Coal Enterprises. *Coal*, **2,** 69-70 (in Chinese)

[14] Wang Dong (2008) *Study on decision support system of the comprehensive budget management of enterprise.* Dissertation of Tianjin University Ph.D. (in Chinese)

Author

Wang Bang-Jun, born in 1975

Current position, grades: postdoctoral, association profession in school of management, China University of Mining & Technology. Members of China Soft Science Research and International Association for Chinese Management Research (IACMR),
Scientific interest: mining construction, information and coal industrial organization

Method for defining multiple homogeneous activities in distributed workflow management system

Yinzhou Zhu*, Hui Yang, Baolin Yin

National Laboratory Software Development Environment, Beihang University, Beijing 100191, China

Received 1 January 2014, www.tsi.lv

Abstract

In current process-oriented software systems, most of the processes have large number of parallel activities, which are homogeneous. These parallel activities are often used in the split-merge workflow structure and make the workflow model too complex to manage, as in the traditional workflow management systems each activity has to be defined respectively and bind to one resource. In this paper, we explore a novel method to define the distributed workflow model, which replaces the multiple homogeneous parallel activities with a batch-activity node to simplify the workflow model. An architecture is designed based on this method, which involves the model of organization structure, resource allocation and the sub-workflow. This architecture allows one batch-activity node bind to multiple resources, which are distributed, over a wide geographic area. Real-world scenarios, which are built and implemented based on this architecture, are shown to prove the effectiveness and usefulness of the method.

Keywords: workflow management, business process, distributed systems, resource allocation, multiple-instances pattern

1. Introduction

In the past decade, workflow management technology has played an important role in the fields of business process management. More and more enterprises consolidate their project implementation into a workflow management system [1, 2]. In the traditional workflow management systems, the tasks on one activity are often allocated to one resource to execute. This model meets the challenge, as the modern business processes have lots of parallel activities which are homogeneous and are distributed over a wide geographic area. These parallel activities make the workflow model very complex, and increase the difficulty of business process management, especially when the process is deployed on a distributed mobile network. In previous studies, this problem is generally attributed to the workflow patterns involving multiple instances [3]. However, the existing multiple-instances patterns are focus on the run-time mechanism, and they can only support the simple split-merge workflow structures. However, the workflow structures in the real world are far more complicated.

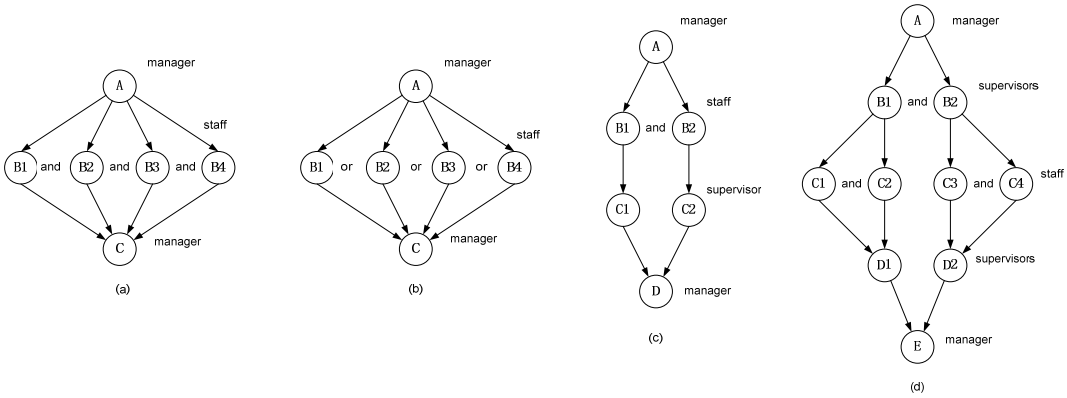


FIGURE 1. The diagram of the split-merge workflow structure

The split-merge workflow structure often contains a lot of homogeneous activities, which have the similar tasks and are allocated to the resources that belong to the same role. Some typical split-merge workflow structures are shown in Figure 1. The structures (a) and (b) contain the classical workflow patterns of AND-split and XOR-split respectively. In above structures, the activities with the B prefix are homogeneous activities. The structures (c)

* Corresponding author.
E-mail: zhuyinzhou@nlsde.buaa.edu.cn

and (d) contain more complex workflow patterns, which are extended from above basic workflow patterns. These structures are more common in the real-world workflows, and we define them as the multi-level split-merge workflow structures. In these structures, the activities with the B prefix are the first-level homogeneous activities; the C and D activities are the homogeneous activities in the deeper level. In real-world workflows, the number of homogeneous activities can be quite large. Therefore, a method that can simplify the description of the multiple homogeneous activities in distributed environment is necessary.

Along the lines presented in this paper, we propose a description model of multiple homogeneous activities (MHA) that can simplify the description of the complicated multi-level split-merge workflow structure. This model is based on a multiple resource allocation model, which use one "MHA" node to represent multiple activities, and a sub-workflow model, which can extend a "MHA" node to a deeper level workflow. An architecture based on this model is designed and implemented in the EasyWork system. The EasyWork system is a distributed workflow management system that developed by our group [4, 5]. It is based on the persistent messaging mechanism, decentralized distributed workflow model, and the universal data bus mechanism.

This paper focuses on the description model and the runtime support mechanism of the multiple homogeneous activities. And many real-world workflow scenarios developed based on EasyWork system is discussed to prove the effectiveness and usefulness of the architecture.

2. Related Works

The distributed workflow management technology is mainly applied in scientific computation environments [6]. With the advent of rapid evolution of the business process in large enterprises, however, distributed workflow management is attracting much attention in business process management field these days [1].

The problem of the defining multiple homogeneous activities in the workflow model is generally attributed to the workflow patterns involving multiple instances in previous studies. However, as these multiple-instances patterns covered a lot of ground in a short space, they were only discussed and implemented at a junior level. The workflow management systems that support these patterns were very rare, let alone the distributed workflow management systems [3].

In recent works on distributed workflow management systems, Muthusamy et al developed a flexible and distributed platform to develop, execute, and monitor business process [7, 8]. This platform supports service discovery and composition among multiple resources that offer the same functionality, but cannot simplify the description of the multiple homogeneous activities. Khalaf and Leymann present a BPEL fragmentation covering data and explicit control dependencies, and an approach to handle fragmenting loops and scopes [9]. Hamann et al present a migration data meta-model for business processes with the ability for runtime migration, which enhance. The flexibility of the distribution of the ad-hoc workflow [10]. The workflow description models in these researches are extended from the BPEL, and cannot descript the multiple homogeneous activities in simple forms.

Besides, most of common business process management systems, such as Staffware, WebSphere, FLOWer, and COSA, are able to support some multiple-instances patterns though the extend mechanisms, such as "bundle model", or "dynamic parallel process management table" [3, 11, 12]. However, their description method is too complex for normal people, and can not support the complicated workflow model that contain multi-level split-merge workflow structures, and most of them can only deployed in a centralized environment.

Our work distinguishes itself from these other approaches by concentrating on a small part of the multiple-instances patterns - multiple homogeneous activities. This help to reduce the model's complexity, make the workflow description simpler and easier to study, and support far more complicated workflow structure than previous patterns.

3. MHA Model

The description model of MHA is composed of three fundamental models: organization structure model, resource allocation model, and sub-workflow model.

3.1. ORGANIZATION STRUCTURE MODEL

The organization structure model, on which the MHA model is based, is derived from EasyWork System. The model consists of four main elements: Department, Workgroup, role and resource. Department and Workgroup are the basic units of the organization structure. They are organized into tree-like hierarchical structures. A department can be only attached to a department, and a workgroup can be only attached to a workgroup. Each department and workgroup can have many roles and resources. The difference between Department and Workgroup is that one resource can and must attached to only one department while it could attached to many workgroups. The role is used to classify the resources by their position or job, such as Manager, Staff, and so on. Each resource could play one or more roles. For example, staff A is the manager of the department X, while at the same time doubled as the manager of the department Y. The diagram of the organization structure model can be seen in Figure 2.

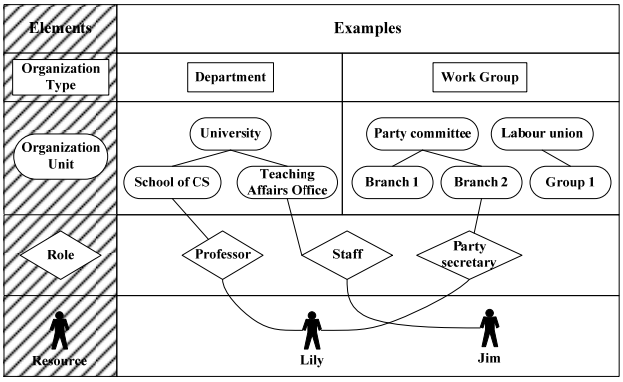


FIGURE 2. The diagram of the organization structure model in EasyWork System

3.2. RESOURCE ALLOCATION MODEL

The resource allocation model is the main part of the MHA model, because it explains the MHA node. A MHA node is an enhanced activity node, which can represent one or more homogeneous activities in order to simply the complexity of the workflow model. The resource allocation description on a MHA node is based on the Organization Unit, which includes departments and workgroups in the organization structure model. The format of the resource allocation description is like this: **quantifier** (*spec_role*) **in/of** (*org_unit*)

In this line, **quantifier** is a word that can be chose from ALL, ANY and THE. "ALL" means the tasks on the MHA node should be distributed to all the resources which under the *org_unit* and play the *spec_role*; "ANY" means the tasks should be allocated to any of the given resources; "The" means the tasks should be allocated to the only one resource that is given. The *spec_role* represents the role of the resources to which the tasks should be allocated. Besides the user-defined roles, such as "manager", "staff", there are three built-in roles: Dept, Group and Node. The Dept represents the departments, the Group represents the workgroups, and the Node represents all the resources under the *org_unit*. The *org_unit* represents the Organization Unit that the resources should be attached to. Besides the user-defined organization unit, there are two built-in unit "this_domain" or "this_node", which are often used in sub-workflow and represents the default organization unit in the current workflow.

TABLE 1. Description examples of the resource allocation

Examples
all (student) **in** (classA)
any (teacher) **in** (schoolB)
the (headmaster) **of** (schoolC)
all (Node) **in** (this_domain)

The Table 1 gives some description examples of the resource allocation on MHA nodes: It shows the understandability of the description model.

3.3. SUB-WORKFLOW MODEL

The sub-workflow here means the workflow that replaces the functions of an activity in the upper level workflow model. In the MHA model, sub-workflows are used to extend the functionality of a MHA node. It helps the MHA model support the multi-level split-merge workflow structures. The model of the sub-workflow is similar as the common workflow model, which is represented as a directed graph that consists of many activity nodes and transition paths. The difference from the common workflow model is that the start activity node and the finish activity node must be allocated to the same resource as the activity from which this sub-workflow extended.

In traditional workflow management systems, the sub-workflows are often used to increase the reusability of the process fragment. More than that, the sub-workflows in the MHA model are focus on simplifying the description of the multi-level split-merge workflow structure. For example, the workflow structures (c) and (d) in Figure 1 are too complex to define only based on MHA nodes. To resolve this problem, we extend the MHA node though a sub-workflow, which can be seen in Figure 3. The diagrams (c') and (d') in this figure show the description models of the workflow structures (c) and (d) in Figure 1 respectively.

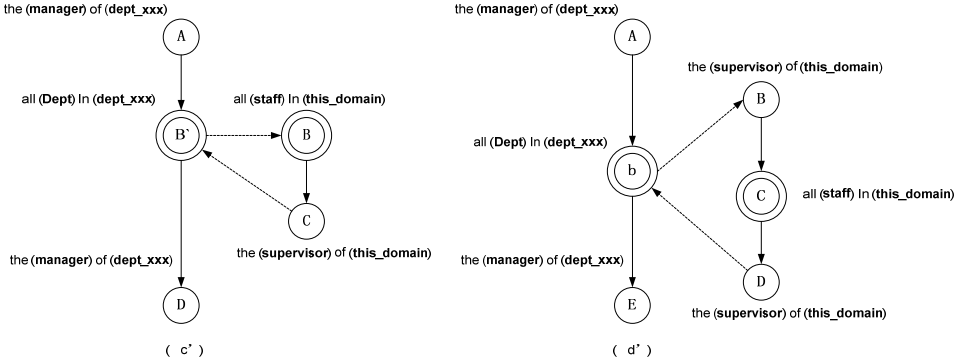


FIGURE 3. The diagram of the sub-workflow model

In the sub-workflow model, which is, extend from a MHA node, the tasks of activities can be allocated to the resources that play the role under the default organization. The default organization is not an absolute organization, but the each organization defined in the MHA node. This method opens the possibility of defining the workflow model through a recursive process, so that simplifies the description of the multi-level split-merge workflow structure.

4. Execution Mechanism

The basic execution mechanism is provided by EasyWork distributed workflow management system. Besides, we extend the mechanisms to support the XOR-split workflow pattern and the data convergence in MHA synchronizing merge workflow pattern.

4.1. ARCHITECTURE OF EASYWORK SYSTEM

The architecture of the EasyWork system, which is shown in Figure 4, aims to define, execute and monitor the workflows for the cross-regional enterprises, which often have several highly autonomous subsidiaries. The network architecture of the EasyWork system is a hybrid structure, which is based on the peer to peer network and the client/server framework. There are two kinds of nodes in the EasyWrok network: EasyWork Server node and the EasyWork Client node. The EasyWork Servers are the basic nodes, which are used to store and dispatch the distributed workflow instances. The relationships between EasyWork Servers are symmetrical, while the EasyWork Clients are client nodes of the EasyWork Server. The EasyWork Platform is installed on every EasyWork Server, and offer the access interface that allows users to get and do their job in remote EasyWork Clients though a standard web browser.

EasyWork Process Definition Server is a kind of EasyWork Server on which the workflow definition tool is installed. The workflow definition tool is used by workflow administrators to define, compile and deploy the distributed workflow. After a workflow model has been designed in the Process Definition Server, it will be split into several segments by activities, and be compiled to configuration files. These configuration files will then be distributed to EasyWork Servers, which are defined as the computing resources of the activities.

The EasyWork Platform on EasyWork Server is composed of three main parts: the workflow engine, the user task manager and the application framework. The workflow engine is used to receive, store, and dispatch the workflow instances, and invoke the application to process the tasks of the activities. The user task manager offers the access interface to the workflow system, shows the task-lists, and communicates with the workflow engine to process user commands. The application framework is a set of applications, which are created, based on the components with application-level granularity [4, 13, 14]. These applications, which are invoked by the workflow engine to process the tasks of the activities, are executable programs, such as the executable binary files of the operational systems of Microsoft Windows or UNIX, or the web pages that could be interpreted by http server or client browser.

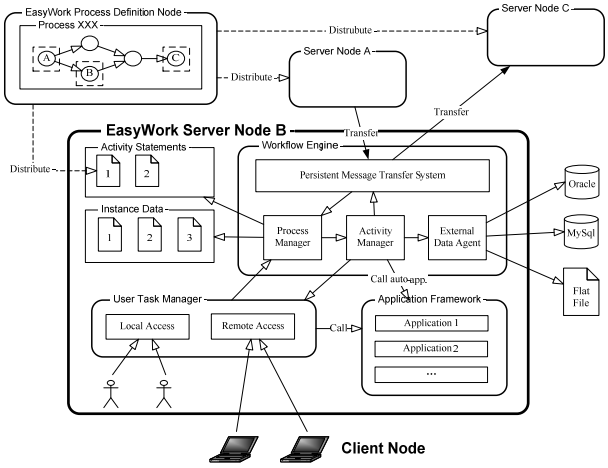


FIGURE 4. The architecture diagram of the EasyWork system

4.2. MECHANISM FOR XOR-SPLIT PATTERN

The workflow models that include the MHA nodes or the sub-workflows is converted to the normal workflow model which only contains the simple activities before the deployment. Then the tasks of the activities are distributed to the resources that would execute them. To support AND-split pattern in the EasyWork system is easy and direct, but to support XOR-split pattern is difficult, because the latter involves the dynamic work allocation in the runtime phase in the distributed environment.

To resolve this problem, a program named “XS” is developed and is added to the end of task queue of every XOR-split activity. The XS knows the number of the activities that can be select as the successor by the XOR-split activity, and the identifier and the weight of each of these activities. When XS is executed, it select the proper activity as the successor based on the serial number of the current workflow instance, and then route the workflow instance. The successor selection can be made based on round-robin, weight-based or random algorithm in current system, and can be extended easily by modifying the XS program.

4.3. MECHANISM FOR DATA CONVERGENCE

Data convergence is a difficult problem when merging multiple parallel workflow instances in AND-split pattern. In the traditional workflow management systems, the workflow data of the parallel workflow instances are often imported into the database individually, and then are fetched together by the synchronization activity. However, this approach increases the coupling between

the workflow applications and the database, and is inappropriate in the distributed environment especially.

To solve this problem, our system adopts a specialized model to describe the data convergence in the synchronization activity, and provide a specialized data type – "collect data". When the workflow instance go through the MHA node which split the instance into multiple parallel workflow instances and merge them after finish the tasks, EasyWork system would gather all collect data and combine them into an array.

The format of the declaration of the collect data is like this: **Define_collect_var** (*collect_array*)

In this line, *collect_array* is the workflow data which would be used to store the combined data of the multiple parallel workflow instances. The data model of the *collect_array* after the data convergence is a relative table, which is shown in following table.

TABLE 2. The model of the collect data

Collect Array
{{node1_id, data1, ...}, {node2_id, data2, ...}, { … }, {nodeN_id, dataN, ...}}

In Table 2, the variables from "node1_id" to "nodeN_id" store the identifier of the multiple activities represented by MHA; the variables with the "data" prefix store the values of the collect data of multiple parallel workflow instances. The data model of the variables that are combined can be simple data, list or relative table. A diagram of the data convergence model when using MHA nodes can be seen in Figure 5. The model at the left of the figure is a normal AND-split workflow, and the model at the right side is the same model but adopts the MHA node.

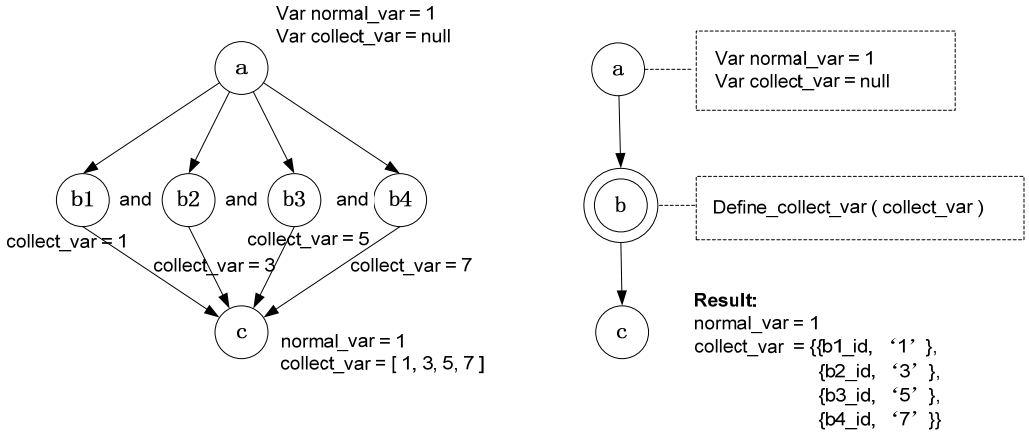


FIGURE 5. The diagram of the data convergence model when using MHA nodes

5. Scenarios

The real world scenarios, which implemented based on EasyWork system, are discussed in this section. As shown in Table 3, the scenarios are grouped into three classes:

(1) Office automation (OA) applications, which are the collaboration systems based on the workflow. The OA system is a type of classic workflow system, and plays an important role in paperless office. The OA systems are widely used almost in all fields without the limit of the industry or geographic region. The key element of the OA is the form, which often is easy to be customized based on the web components in the EasyWork system.

(2) Enterprise resource planning (ERP) system, which is a more integrated platform, compared to OA systems, used to manage various kinds of information in the enterprise. From the model layer, such as the data model and business logic, to the view layer, such as the view of the decision support, all kinds of the information in the enterprise are integrated together into a strict architecture. The design of the data model and process in ERP system are more professional than other information management systems, and are more difficult to be implemented.

(3) Report generation systems, which are the process-oriented systems used to collect and summarize the information, which is distributed in different areas, and generate the Summary Report finally. This kind of application is very common in different areas, and is not easy to implement for the complex computation and distributed architecture. The real-world report generation scenario implemented based on EasyWork system is the Freshwater Quality Monitoring (FQM) which is a process required by the National Oceanic Administration of China to investigate the freshwater quality in the coastal areas and provide a summary report every year.

TABLE 3. Real world workflows implemented by EasyWork system

Class		Total number	AS5	XS5	AS20	XS20	Depth
OA	Document Circular	3	2	3			
	Administrative Approval	6	2	4			3
	Personnel Management	9	6	3	1		1
	Financial Management	3	2	2			2
	Product Management	6	1	6			
	Customer Service	5	1			3	
ERP	Data Maintenance	15	6	10	3	4	7
	Purchase Processes	6	2	4		1	2
	Sale Processes	7	2	4	2	1	3
	Bank Processes	4	2	3			
	Inventory Management	4	2	4			
	Production Processes	5	1	4			
	Report Query	18	3	12	1	1	
Other Processes		4	1	3	3	1	1
	Summary	95	33	62	10	11	19
		100%	*34.74%*	*65.26%*	*10.53%*	*11.58%*	*20.00%*

In the above scenarios, there are 95 workflows implemented by EasyWork system. The details of the workflows are shown in Table 1. The "AS5" and "AS20" represent the workflows that involves AND-split workflow pattern and the max number of the multiple heterogeneous activities is more than five and twenty respectively. In the same way, the "XS5" and "XS20" represent the workflows that involves XOR-split workflow pattern and the max number of the multiple heterogeneous activities is more than five and twenty respectively. The "depth" means the total number of levels of the multi-level split-merge workflow. The Table 1 shows that, nearly half of the business workflows that involve the AND-split workflow pattern are suggested to adopt the MHA model, and 20% of the them have to adopt the MHA model; nearly 4/5 of the business workflows that involves the XOR-split workflow pattern are suggested to adopt the MHA model, and 15% of them have to adopt the MHA model. Although these statistics are only based on the experience gathered in our work history, they still explain the seriousness of the MHA problem to some extent.

5.1. CASE STUDY

To illustrate the effectiveness of the MHA description model, a real-world scenario which described by the EasyWork system is shown in this section. The workflow in this scenario is used by a state-owned enterprise to reimburse the project expenses spent during the last year. The diagram of the organization structure of the enterprise is shown in Figure 6.

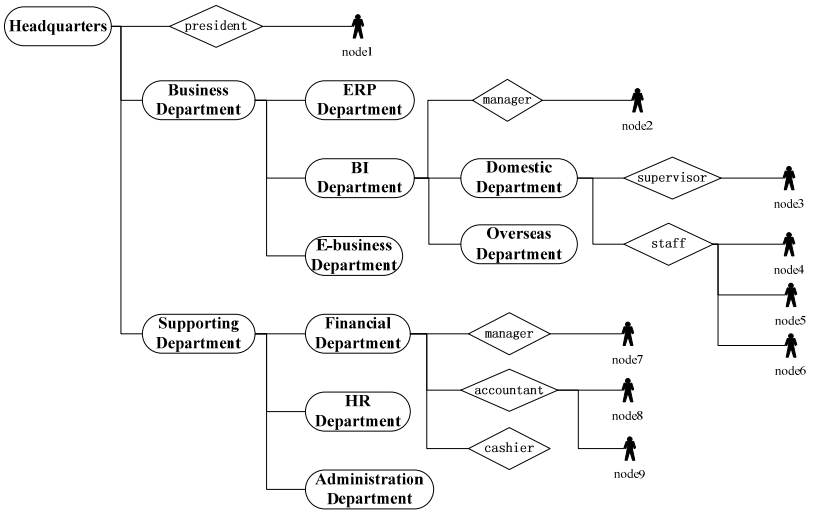


FIGURE 6. The diagram of the enterprise organization structure

The reimbursement workflow is started by the manager of the financial department, and then the reimbursement instructions are forwarded to every business department. After every supervisor filled out the expenses claim form, the forms are grouped and submit to their managers to approve. If the application is not

approved, it would be sent back and be filled again. After the verification of the department manager, all expenses claim forms are gather together and transfer to the accountant of the financial department to do the second check and create the expenses report. Then the report is checked by financial manager and is delivered to the cashier to do the banking business. The model of the workflow can be seen in Figure 7.

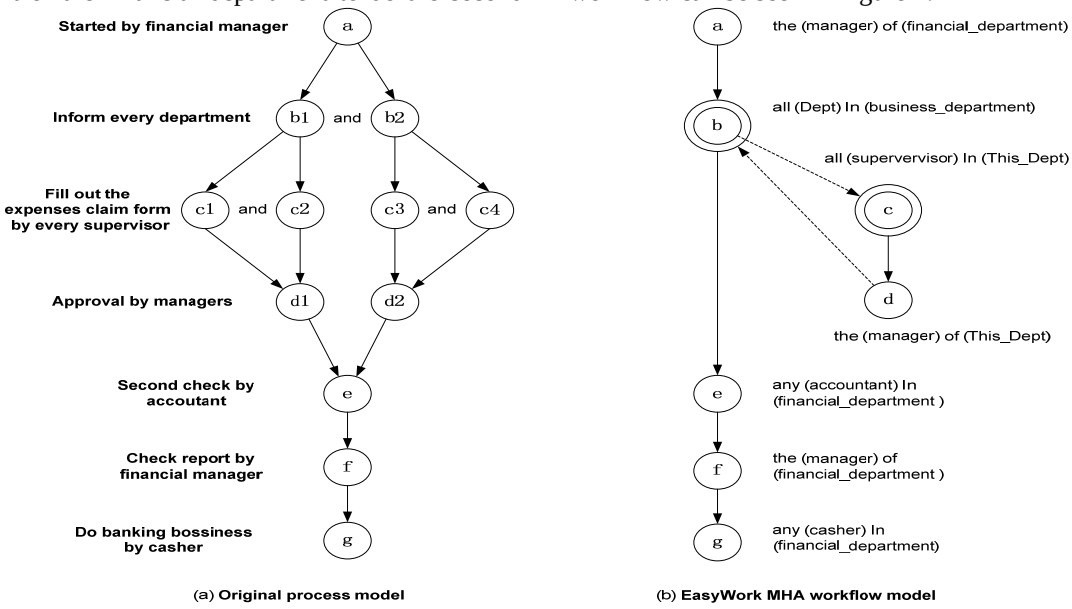


FIGURE 7. The model of the annual reimbursement workflow for project expenses

This reimbursement workflow is a complex workflow that has two-level split-merge workflow structure, but it is very common in the real-world scenarios. In the enterprise that was described above, the actual number of the departments attached to the business department is 16. Moreover, each of these departments has more than 10 sub-departments. This means in second level of the workflow model, there are about 200 heterogeneous activities. It is very difficult to mange this complex workflow model by traditional workflow management system. However, EasyWork system can play a maximum efficacy in this situation. Two MHA nodes and a simple sub-workflow are enough to describe this model. Besides, the descriptions of resource allocation are short and easy to understand. It shows the high efficiency and the usability of the MHA description model.

5.2. SYSTEM COMPARISON

This section discusses the comparison between EasyWork system, which implements the MHA description model and the other mainstream workflow management systems. The workflow management systems to which we compared are Staffware, COSA, FLOWer, WebSphere MQ Workflow and SAP/R3. The comparisons can be seen in Table 4. Some part of the information in this table is derived from [3].

TABLE 4. Comparisons between EasyWork system and other WFMS

Systems	AND-split MHA	XOR-split MHA	Multi-level MHA	Supporting distributed workflow	MHA data convergence mechanism	MHA task auto distribution
Staffware 9	-	+	-	-	-	-
COSA 4.2	-	-	+ / -	+	-	+ / -
FLOWer 3	+	+	-	-	-	-
Meteor	-	-	-	+	-	-
WebSphere MQ Workflow 3.3.4	-	+	-	-	-	-
SAP R3	+ / -	+	-	-	-	-
EasyWork	+	+	+	+	+	+

The table shows that the Staffware and Websphere MQ Workflow are only support the basic XOR-split MHA description. The SAP/R3 and FLOWer can support basic AND-split MHA description but cannot support multi-level MHA description. The COSA support the multi-level workflow and task auto distribution, but it cannot support MHA description well. In the last, all these traditional workflow management system do not support MHA data convergence. Compare to these system, EasyWork support all these functionalities. Besides, the description rules in EasyWork system are very simple and easy to learn and use.

6. Conclusions

In this research, we propose a workflow management framework for multiple heterogeneous activities. This framework supports the description and execution of the MHA nodes, and increases the efficiency of workflow model management. Compared to other mainstream workflow management systems, this framework offers much fuller support on MHA management. The MHA description model involves organization structure, resource allocation model, and sub-workflow description model, and offers two key mechanisms for the automatic task distribution in the XOR-split pattern and the data convergence in AND-split pattern. In the future, the relationship between resource allocation model and the sub-workflow description model will be enhanced in order to increase the flexibility on defining the resource allocation in multi-level workflows.

References

[1] Van der Aalst W. M. P. (2013) Business Process Management: A Comprehensive Survey. *ISRN Software Engineering*, **2013**
[2] Sinur J., Hill J. B. (2010) *Magic Quadrant for business process management suites.* Gartner RAS Core research note, 1-24
[3] Van der Aalst W. M. P., Ter Hofstede A. H., Kiepuszewski B., Barros A. P. (2003) *Distributed and parallel databases*, **14**(1), 5-51
[4] Yinzhou Zhu, Baolin Yin (2013) *Journal of U-& E-Service, Science & Technology*, **6**(4)
[5] Yinzhou Zhu, Baolin Yin 2013 *Journal of Computational Information Systems*, **9**(20), 8297-306
[6] Deelman E., Gannon D., Shields M., Taylor I. (2009) *Future Generation Computer Systems*, **25**(5), 528-540
[7] Muthusamy V. (2011) *Flexible Distributed Business Process Management:* University of Toronto.
[8] Ganesan S., Yoon Y., Jacobsen H. A. (2011) NIÑOS take five: the management infrastructure for distributed event-driven workflows. In *Proceedings of the 5th ACM international conference on Distributed event-based system, DEBS*, 11, 195-206
[9] Khalaf R., Leymann F. (2012) *Information Systems*, **37**(6), 593-610
[10] Hamann K., Zaplata S., Lamersdorf W. (2012) Process instance migration: Flexible execution of distributed business processes. In *Software Services and Systems Research-Results and Challenges* (*S-Cube*) (2012) Workshop on European IEEE
[11] Russell N., van der Aalst W. M. P, ter Hofstede A. H., Edmond D. (2005) Workflow resource patterns: Identification, representation and tool support. *Advanced Information Systems Engineering Heidelberg: Springer*
[12] Russell N., ter Hofstede A. H., Edmond D., van der Aalst W. M. P. (2005) *Conceptual Modelling–ER*, **3716,** 353-68
[13] Guojie Jie, Baolin Yin, Qiyang Zhao (2011) *Journal of Software*, **6**(3), 374-85
[14] Guojie Jin, Baolin Yin, Qinyang Zhao (2011) *Computer Integrated Manufacturing Systems*, **17**(2), 425-32

Authors

Yinzhou Zhu

University studies: M.S. in 2006 and currently is a Ph.D. candidate, both in Computer Science and Engineering Department at Beihang university (Beijing)
Scientific interest: workflow management, enterprise application integration and distributed systems

Hui Yang

University studies: Bachelor in 2011 at Xidian University and currently is a M.S. candidate, in Computer Science and Engineering Department at Beihang university (Beijing)
Scientific interest: workflow management, Web component and database

Baolin Yin

Current position, grades: Professor in Computer Science and Engineering Department at Beihang university (Beijing), and is currently head of the State Key Laboratory of Advanced Software Development
University studies: He received his Ph.D. in Artificial Intelligence in the University of Edinburgh (1984)
Scientific interest: His research activity is concerned with distributed systems, workflow management, image identification, and computer security

Empirical research on existing quantity of small and medium-sized enterprises in China, based on system dynamics

Wang Mingpeng[1] *, Sun Qiming[2]

[1] *School of Economics and Management, Beijing University of Posts and Telecommunications, Room 508, Teaching Building 1, Beijing University of Post and Telecommunications, No.10 Xitucheng Road, Haidian District, Beijing, China*

[2] *School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing University of Post and Telecommunications, No.10 Xitucheng Road, Haidian District, Beijing, China*

Received 1 January 2014, www.tsi.lv

Abstract

A system model for the existing quantity of small and medium-sized enterprises is built in this document, by establishing relationship equations with study on relationships among more than 30 variables such as total enterprise quantity, establishment rate of new enterprises, level of human resources, level of technical innovations, index of resource dependence, etc. Moderate breakthroughs are made on the mathematical methodology, such as the method of education years to calculate the level of human resources, the method of resource dependency evaluation for the resource dependency index. However, certain corrections are made for adaptation to the study. The innovative concept of establishment rate of new small and medium-sized enterprises is created in modelling and correlated with level of technical innovations, level of human resources and resource dependency index through relationship functions. The purpose thereof is to explore mechanisms where and extents to which influence factors make impact on the existing quantity of small and medium-sized enterprises. Finally, emulation prediction for the system model is made with the emulator Vensim and the error analysis on comparison between emulation and historical data is performed. It is found that the agreement with historical data is good and the error is acceptable.

Keywords: existing quantity of small and medium-sized enterprises, establishment rate of new small and medium-sized enterprises, system dynamics

1. Introduction

The system model for existing quantity of small and medium-sized enterprises is used for predictive investigation on changing quantity of enterprises as well as for analysis of how and what the impact of influence factors on existing quantity of enterprises would be during a period of time. To make convenient and reasonable system modelling, the influence factors are divided into five subsystems correlated via certain relationships.

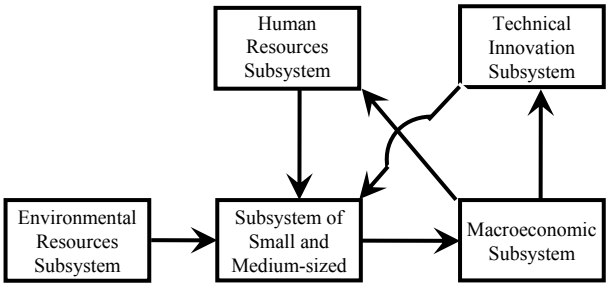


FIGURE 1. Structure of the System Model for Existing Quantity of Small and Medium-sized Enterprises

2. Foundational Model Analysis

2.1. MODELLING PRINCIPLE

2.1.1. Analysis of System Architecture:

The establishment of small and medium-sized enterprises is influenced by multiple factors such as economy, man power, technology, policy and environment. The knowledge flows in the form of technology from the technical innovation subsystem to the subsystem of small and medium-sized enterprises. The effect of such outflow is also presented in education. The quantity of well-educated labors determines the level of human capital required by entrepreneurs and small and medium-sized enterprises, particularly small high-tech enterprises. Such labors correlate the human resources subsystem with the subsystem of small and medium-sized enterprises and flow from the former to the latter. The abundance of environmental resources influences economic structure and economic development centering within a region. The resource deterioration results in poor innovation ability and low entrepreneurial efficiency. Thus, the environmental resources subsystem provides control effect on the subsystem of small and medium-sized enterprises. Normally, more the quantity of small and

* Corresponding author. Tel: +86-138-1002-1390; fax: +86-010-6228-1998;
E-mail: 115636795@qq.com

medium-sized enterprises is, larger the contribution they make to gross domestic products will be. The corresponding taxes will be invested into education and R&D through the government's financial inclination. The funds will then flow from subsystem of small and medium-sized enterprises to macroeconomic subsystem, human resources subsystem and technical innovation subsystem, and finally make a fed impact on the subsystem of small and medium-sized enterprises.

2.1.2. Description of Variables

The human resources subsystem will explain variation of the human capital level with a focus on the level of human resources. The technical innovation subsystem will explain the extent of knowledge outflow with a focus on the level of technical innovations. The environmental resources subsystem will explain dependency of economic development modes on resources in a region with a focus on resource dependency index. The macroeconomic subsystem will explain economic development trends with a focus on gross domestic products. The subsystem of small and medium-sized enterprises will explain variation mechanism of the existing quantity of small and medium-sized enterprises with a focus on total enterprise quantity.

2.2. STRUCTURAL FLOW CHART

2.2.1 Key Feedback Loops

- Total enterprise quantity →+ Gross domestic products→+ Taxes →+ Financial revenue →+ Financial expenditure →+ Investment on R&D →+ Existing R&D quantity →+ Level of technical innovations →+ Establishment rate of new enterprises →+ Quantity of newly established enterprises →+ Total enterprise quantity
- Total enterprises quantity →+ Gross domestic products →+ Taxes →+ Financial revenue →+ Financial expenditure →+ Educational investment →+ Educational investment coefficient →+ Educational coefficient →+ Level of human resources →+ Establishment rate of new enterprises →+ Quantity of newly established enterprises →+ Total enterprise quantity

2.2.2. Flow Positions and Flow Rates

- Flow position: total enterprise quantity, flow rate: quantity of newly established enterprises, quantity of bankrupt enterprises
- Flow position: existing R&D quantity, flow rate: R&D investment, R&D decrease
- Flow position: total population quantity, flow rate: new population, death population.

2.2.3. Emulation Model

Model emulation is made with the emulator Vensim. The structural flow chart is as shown in the Figure 2.

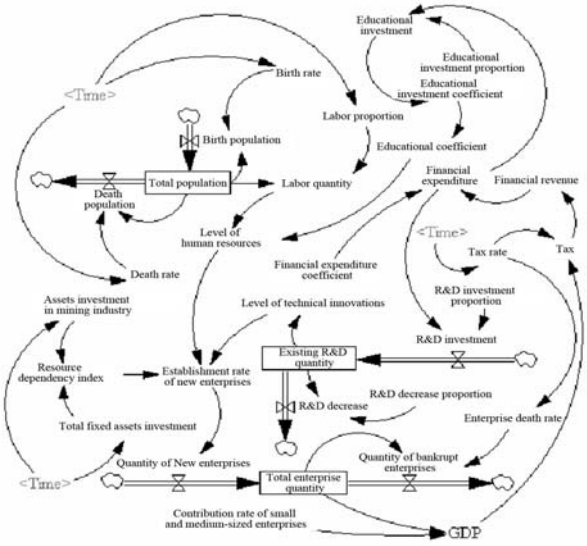


FIGURE 2. Emulation model made with Vensim

3. Establishment of Modelling Equations*

3.1. POPULATION-RELATED VARIABLES

The index linear regression† with the birth rate and natural growth rate of Chinese population during 1999 and 2008 and the time is made according to Malthus' theory of population index growth. PBR_t (birth rate of the population in year t) and PDR_t (death rate of the population in year t) are obtained as below.

$$PBR_t=23.128\cdot e^{-0.0326(t-1986)}/1000,$$

$$PBR_t=23.128\cdot e^{-0.0326(t-1986)}/1000-17.622\cdot e^{-0.0326(t-1986)}/1000$$

It is known from definitions of birth and death rates that:

$$PBR_t=\frac{PB_t}{(POP_{t-1}+POP_t)/2} \qquad PB_t=\frac{2POP_{t-1}}{2-PBR_t} \qquad PD_t=\frac{2POP_{t-1}}{2+PDR_t}$$

$$POP_t=\int_{2000}^{t}[PB(i)-PD(i)]\cdot di+POP_{2000}$$

$$=\int_{2000}^{t}[PB(i)-PD(i)]\cdot di+127627,$$

where POP_t is the total population in year t

3.2. LABOR-RELATED VARIABLES

LR_t (the labor proportion):

$$LR_t=-0.0001\cdot(t-1999)^2+0.003\cdot(t-1999)+0.5739,$$

$$L_t=POP_t\cdot LR_t,$$

* All the data used to establish the equation are totally from annual China Statistical Yearbook of National Bureau of Statistics.

† All the linear regressions mentioned in this article are analysed through IBM SPSS Statistics 19.0.0, specifically not repeat.

HRL_t (level of human resources in year t):

$$HRL_t=HR_t/HR_{2000}=EC_t\cdot L_t/73991.98.$$

3.3. EDUCATION-RELATED VARIABLES

EIR (educational investment proportion):

$$EIR=\frac{\sum_{i=2000}^{2009} EIR_i}{2009-2000+1}=12.5168794\%$$

EI_t (educational investment): $EI_t=FE_t\cdot EIR$, where FE_t is the financial expenditure in year t.

The methodology for calculation of the human capital level can be divided into future proceeds method, accumulative cost method, education years method and income differentiation method. The education years method is used for human capital calculation in this model because of limited data acquisition and method unsuitability. Angus Maddison's weighting criteria are used for the effect of academic qualifications on human capital. The weight of tertiary education is 2, and the weight of secondary education is 1.4, and the primary education is 1. The calculation method is as below:

HR_t (total human resources in year t):

$$HR_t=\int_0^{\infty}\alpha\cdot\eta(t,\alpha)\cdot L_t d\alpha=\left[2\cdot\eta(t,2)+1.4\cdot\eta(t,1.4)+1\cdot\eta(t,1)\right]\cdot L_t$$

(Where α is the education extent that is the weight given and $\eta(t,\alpha)$ is the proportion of the population with education extent in the total population in year t)

EL_t (education level in year t):

$$EL_t=2\cdot\eta(t,2)+1.4\cdot\eta(t,1.4)+1\cdot\eta(t,1)$$

EIR_t (educational investment coefficient in year t):

$$EIR_t=EI_t/EI_{2000}=EI_t/1624.1438$$

EC_t (educational coefficient in year t):

$$EC_t=0.021\cdot EIC_t+0.997 \qquad R^2=0.8260$$

3.4. VARIABLES RELATED TO R&D INVESTMENT

The proportion of Chinese R&D investment in the financial expenditure is 34.0% - 44.4% during 2000 and 2009, and its variable properties and data characteristics are similar to those of the educational investment. Thus, the same method is used for calculation.

RNDIR (R&D investment proportion) and $RNDI_t$ ($R\&D$ investment in year t):

$$RNDIR=\frac{\sum_{i=2000}^{2009} RNDIR_i}{2009-2000+1}=3.775725\%$$

$$RNDI_t=FE_t\cdot RNDIR$$

3.5. VARIABLES RELATED TO R&D DECREASE

The R&D will be decreased along with the time. The R&D decrease proportion is the proportion between R&D decrease in a period and R&D quantity in beginning of the period. RNDDR(R&D decrease proportion) and $RNDD_t$(R&D decrease in year t) are obtained with testing of historical experience data, as specified in this document.

$$RNDDR=0.2 \qquad RNDD_t=RND_t\cdot RNDIR$$

3.6. VARIABLES RELATED TO R&D QUANTITY

The R&D quantity in a period is the sum of R&D quantity in the last period and R&D increase in this period (that is the difference between R&D investment and R&D decrease). Thus, RND_i (R&D investment in period i) is:

$$\begin{aligned}RND_i&=RND_{i-1}+RNDI_i-RNDDR\cdot(RND_{i-1}+RNDI_i)\\&=0.8\cdot(RND_{i-1}+RNDI_i)\\&=0.8\cdot[0.8\cdot(RND_{i-2}+RNDI_{i-1})+RNDI_i]\\&\cdots\cdots\\&=0.8^iRNDI_1+0.8^{i-1}RNDI_2+\cdots\cdots+0.8RNDI_i\end{aligned}$$

The R&D quantity in the first period is equal to the decreased R&D investment quantity.

$$RND_1=0.8\cdot RNDI_1$$

$$RND_i=\sum_{p=1}^{i}0.8^{i-p+1}\cdot RNDI_p$$

RND_t (R&D quantity in year t) and IL_t (the level of technical innovations in year t) are:

$$\begin{aligned}RND_t&=\int_{2000}^{t}\left[RNDI(i)-RNDD(i)\right]\cdot di+RND_{2000}\\&=\int_{2000}^{t}\left[RNDI(i)-RNDD(i)\right]\cdot di+1542.807\end{aligned}$$

$$IL_t=RND_t/1542.807$$

3.7. RESOURCE DEPENDENCY INDEX

The resource dependency index is defined as the ratio of fixed assets investment in the mining industry to total fixed assets investment, with Auty's method for dividing of regions with abundant and lean resources as well as reference to the resource dependency measurement method used in China:

$$MDI_t=MFAI_t/FAI_t,$$

FAI_t (total fixed assets investment in year t) and $MFAI_t$ (fixed assets investment in the mining industry in year t) are obtained:

$FAI_t=2086\ (t-1999)^2-3019(t-1999)+11929 \qquad R^2=0.9910$

$MFAI_t=95(t-1999)^2-161(t-1999)+497.84 \qquad R^2=0.9934$

3.8. VARIABLES RELATED TO FINANCIAL

FEC (financial expenditure coefficient) and FE_t (financial expenditure in year t) are:

$FEC=1.018$, $FE_t=FEC\cdot FR_t+1821.299$, $R^2=0.9870$.

There are many particular stipulations about taxation imposed on small and medium-sized enterprises in China. For example, several preferential stipulations involving small and medium-sized enterprises are mentioned in Section 4 "Tax Preferences" in Law of Corporate Income

Tax of People's Republic of China, such that "the corporate income tax will be levied at the tax rate of 20% on small and micro-profitable enterprises that meet conditions, and the corporate income tax will be levied at the tax rate of 15% on high-tech enterprises that should get significant national supports". Therefore, there are different tax rates associated with different industries and small and medium-sized enterprises with different sizes. This study is to investigate the effect of external circumstances on creation, development and death of small and medium-sized enterprises. Thus, it is feasible to find the impact of adjustment to tax rates by considering average tax rates as per the ratio of taxes to GDP.

$$TAX_t=GDP_t \cdot TR_t$$

$$TR_t=0.0051 \cdot (t-1999)+0.1269 \qquad R^2=0.9616$$

$$FR_t=1.167 \cdot TAX_t -1661.441 \qquad R^2=1.0000$$

3.9. VARIABLES RELATED TO BANKRUPT ENTERPRISES

In recent years, there are annually 8% - 10% of small and medium-sized enterprises going bankrupt in China. Most of small and medium-sized enterprises are not in existence for more than 5 years. For such enterprises, the tax level will influence directly their profit rates and dominate the quantity of those going bankrupt. Therefore, the enterprise death rate is introduced between the tax and the quantity of bankrupt enterprises, and is defined as the ratio of bankrupt enterprise quantity to total enterprise quantity. It means that bankrupt enterprise quantity is equal to the product of total enterprise quantity and enterprise death rate. A strong linear relationship between enterprise death rate and tax rate is found from data adjustment to the enterprise death rate, and its value is always kept as about half of the tax rate value. SMEDR t (the death rate of small and medium-sized enterprises in year t) and SMEDt (the quantity of small and medium-sized enterprises going bankrupt) are obtained:

$$SMED_t=SME_t \cdot SMEDR_t,\ SMEDR_t=TR_t/2.$$

3.10. VARIABLES RELATED TO TOTAL ENTERPRISE QUANTITY

The total quantity of small and medium-sized enterprises during 2000 and 2004 is estimated with linear regression of statistics (after 2005) about the quantity of small and medium-sized enterprises that are provided by National Bureau of Statistics and iResearch. A medium-sized variable called the contribution rate of small and medium-sized enterprises is introduced to explain the effect of small and medium-sized enterprises on GDP.

SMEC (contribution rate of small and medium-sized enterprises), GDPt (gross domestic products in year t) and SMEt (total quantity of small and medium-sized enterprises in year t) are obtained:

$$SMEC=99.120,$$

$$GDP_t=99.120 \cdot SME_t -73824.273,\ R^2=0.9840.$$

$$SME_t=\int_{2000}^{t}[SMEB(i)-SMED(i)] \cdot di + SME_{2000}$$

$$=\int_{2000}^{t}[SMEB(i)-SMED(i)] \cdot di + 1693.313$$

3.11. POPULATION-RELATED VARIABLES

A medium-sized variable called the establishment rate of new enterprises is introduced to explain the effects of these three factors on new enterprise quantity. The establishment rate of new enterprises is defined as the equivalent value of new enterprise quantity in the present period to new enterprise quantity in the first period (the first period in this model is the year of 2000) as a standard. SMEBt (quantity of small and medium-sized enterprises newly established) is obtained:

$$SMEB_t=SMEBR_t/280.211,$$

where SME_{ts} and SME_{te} are total quantity in beginning of the period and total quantity at end of the period)

$$SME_{te}=SME_t,\ SMEDR_t=TR_t/2,$$

$$SME_{ts}=SME_{te}/(1- SMEDR_t),$$

$$SMEB_t=SME_{ts}-SME_{te}, \qquad SMEBR_t=SMEB_t/ SMEB_{2000}\ .$$

4. Inspection of Model Reliability

4.1. STRUCTURAL INSPECTION

The reasonability of model structure determines whether emulation can be made in the expected direction and whether variation relationships among variables conform to basic principles and practical experiences. Key parts of this model are obtained with the analysis of numerous historical literatures and actual situation about establishment and development of small and medium-sized enterprises in China. The extent to which influence factors make impact on the system has been fit from actual situation in years. Therefore, the structure of this model is basically reasonable and effective in terms of basic principles or practical experiences.

4.2. PARAMETER INSPECTION

4.2.1. Dimensional Inspection

The model can be inspected with the function of "units check" in Vensim. There would be error prompts if the variable dimensions could not be kept consistent. This model has been checked to be correct, and this means that the variable units are consistent.

4.2.2. Parameter Inspection

The results of model emulation vary with parameters. The properness of a parameter value (with positive or negative sign) can be indicated by means of goodness of fit between variables. Strict regression analysis has been used to establish the relationship equations between model variables. Thus, regression equations in the model

have a good fit. There are strong significance correlation and significance from a statistic view, so it is believed that model parameters are reasonable and effective.

Parameter intervals in the emulation cycle are inspected. For example, the interval inspection is made for tax rate, labour proportion, financial expenditure coefficient, etc. It is found that parameter intervals are within a reasonable range. Detailed inspection processes are not mentioned any more.

4.3. HISTORICAL INSPECTION

The error analysis is performed for part of important variables in the model, such as total enterprise quantity, GDP, financial expenditure, total population, tax, fixed assets investment, educational investment, R&D investment, financial investment, fixed assets investment in the mining industry, new enterprise quantity, bankrupt enterprise quantity, etc.

4.3.1. Relative Error

The ratio of the difference between emulation data and real data to the real data is used as the relative error of variable emulation results. In the equation, e_{it} represent the relative error of emulation data to real data that are associated with variable i in year t, and x_{it} represents historical data associated with variable i in year t, and y_{it} represents emulation data, and i(i=1,2,3…) represents variable i and t represents year t: $e_{it} = \frac{|x_{it} - y_{it}|}{x_{it}}$

There are no absolute inspection criteria in theoretical studies with the system dynamics. However, it is described in the book called Econometrics written by Li Zinai that emulation results could be considered as having good predictability if the value of relative error e_it for more than 70% of variables is less than 5% and the relative error for all variables is within 10%4. Calculation of the relative error is performed with use of far variables in the feedback loop as good as possible in order to reflect model reliability, because most of close variables have a strong explanatory relationship with each other. It can be seen from the above table that the errors of important variables observed through emulation such as total enterprise quantity, GDP and total population are within 10%, and particularly the error of total enterprise quantity is within 2%. Thus, relative errors in the model are small (see Table 1).

TABLE 1. Emulation and Real Data and Relative Errors

	Total enterprise quantity			GDP			Total population		
Year	**Emulation**	**Real**	**Error**	**Emulation**	**Real**	**Error**	**Emulation**	**Real**	**Error**
2005	2793	2836	1.50%	198067	184937.4	7.10%	130477	130756	0.21%
2006	3093	3152	1.87%	227770	216314.4	5.30%	131085	131448	0.28%
2007	3435	3453	0.51%	261681	265810.3	1.55%	131651	132129	0.36%
2008	3827	3850	0.58%	300577	314045.4	4.29%	132179	132802	0.47%
2009	4279	4292	0.29%	345350	340506.9	1.42%	132669	133474	0.60%

4.3.2. Mean Square Percentage Error

The mean square index for the relative error of variables in periods is used to calculate the mean square percentage error of variables (see Table 2). In the equation, e_i represents the mean square percentage error of variable i, and e_it represents the relative error of emulation data to real data that are associated with variable i in year t, and i (i = 1, 2, 3…) represents variable i, and t (t = 2000, 2001…2020) represents year t: $e_i = \sqrt{\sum_1^n e_{it}^2 / n}$.

TABLE 2. Mean Square Percentage Error

Variable	Error
Total enterprise quantity	**0.89%**
GDP	**2.62%**
Financial expenditure	**4.80%**
Total population	**0.29%**
Tax	**4.82%**
Fixed assets	**18.07%**
Educational investment	**3.72%**
R&D investment	**4.81%**
Financial revenue	**4.62%**
Mining industry	**3.57%**
New enterprises	**3.65%**
Bankrupt enterprises	**2.78%**

Similarly, emulation results could be considered as having good predictability if the value of mean square percentage error e_i for more than 70% of variables is less than 5% and the mean square percentage error for all variables is within 10%.

It can be seen from the above table that mean square percentage errors of all variables are within 5%, and only e_i for fixed assets investment is 18.07%.

5. Model emulation results and conclusions

It is expected that the quantity of small and medium-sized enterprises in China will be 181,129,000 by the year of 2020, which is increased by 138,209,000 from 42,920,000 in 2009 and at the composite annual growth rate of 12.7485%. GDP will be RMB 171.653 trillion by the year of 2020, which is increased by RMB 137.6024 trillion from RMB 34.0506 trillion in 2009 and at the composite annual growth rate of 14.4311%.

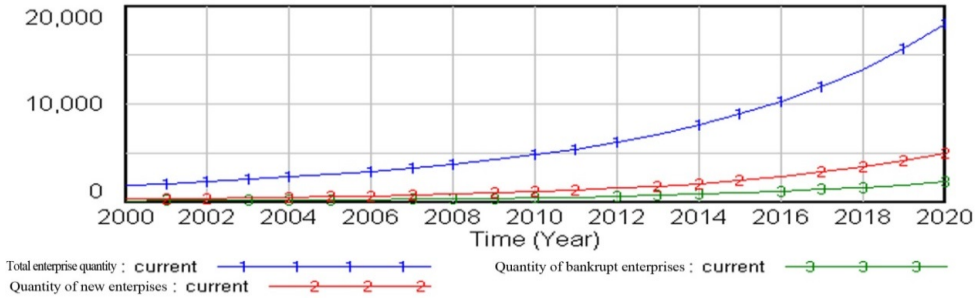


FIGURE 3. The Emulation Result for Quantity of Small and Medium-sized Enterprises

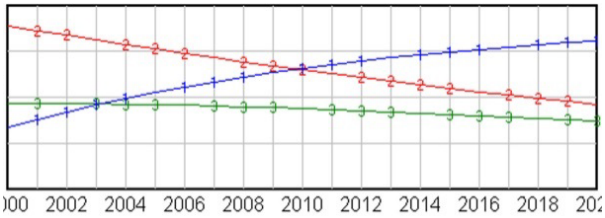


FIGURE 4. The Emulation Result for Population Quantity(1:Total, 2:Birth, 3:Death)

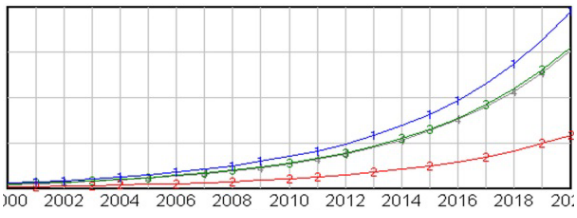


FIGURE 5. The Emulation Result for Finance(1:Eduacational, 2:R&D, 3:Expenditure, 4:Revenue)

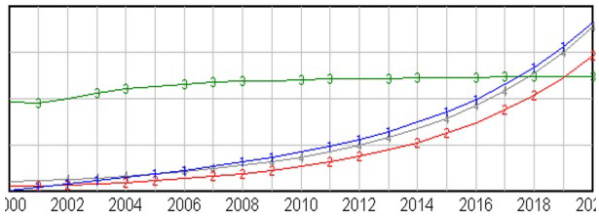


FIGURE 6. The Emulation Result for the Establishment Rate of Enterprise(1:HR, 2:R&D, 3:Dependency, 4:SMEBR)

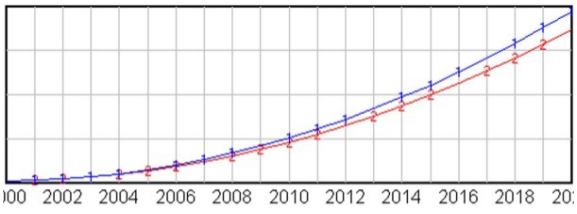


FIGURE 7. The Emulation Result for Resource Dependency(1:Asserts investment in mining, 2:Total)

The quantities of birth population and death population are decreasing annually, but the decrease of birth population is obviously larger than the quantity of death population. Thus, the total population is still increasing but at a growth rate that slows down gradually.

The establishment rate of new enterprises increases with constantly improved levels of human resources and technical innovations, but the resource dependency index is basically kept unchanged. The level of human resources is influenced in its way of increasing by slowdown of the population growth, and this causes slowdown of the growth of new small and medium-sized enterprises. The quantity of R&D investment is then influenced indirectly through the feedback loop, leading to a slight prevention effect. This can be seen from distance variation of the curve representing levels of technical innovations and human resources. The growth of educational investment also slows down due to the effect of level of human resources, but not significantly.

Taxes keep growing synchronously with GDP. Tax rates are increasing but to a small extent from a statistic view, so it is believed that model parameters are reasonable and effective.

References

[1] Peiwen Bo (2010) The Human Capital Level of Chinese Labour Force Estimation: 1989-2007, *Education and Economy*, **2**

[2] Duran-Encalada J. A., Paucar-Caceres A. (2009) *Systemic Practice and Action Research*, **22**(2), 77-99

[3] Dongjie Guan, Weijun Gao, Weici Su, Haifeng Li, Kazunori Hokao (2006) *Ecological Indicators*, **11**(2011), 1333-44

[4] Zinai Li (2000) *Econometrics*, Higher Education Press, Beijing

[5] Yongguang Zhong, Xiaojing Jia, Xu Li (2009) *System Dynamics*, Beijing: Science Press

[6] Dollinger M. J. (1999) *Entrepreneurship: Strategies and Resources*, New Jersey: Prentice Hall

[7] *National Bureau of Statistics of China, Annual China Statistical Yearbook (1990-2010)*, China Statistics Press

[8] Weihua Liu, Yong Zhang (October 21, 2009) The Development Strategies Approach to China's Small and Medium-sized Enterprises. *Proceedings of the 16th International Conference on Industrial Engineering and Engineering Management*, Beijing, China

[9] Zhaoliang Wang (May 13, 2011) The Financial Strategy Research for SME based on the Theory of Corporate Life Cycle. *Proceedings of the 2011 International Conference on Business Management and Electronic Information*, Guangzhou, China

[10] Choi Y. R., Phan P. H. (October 18, 2004) Dynamics of New Firm Formation: Impacts of Economic and Technology Policy. *Proceedings of the 2004 IEEE International Engineering Management Conference on Innovation and Entrepreneurship for Sustainable Development*, Singapore

Authors

Wang Mingpeng, born 1983.02.11, China

University studies: M.Sc. in Business Administration (2008) and PhD in Management Science and Engineering (2011) from Beijing University of Posts and Telecommunications.
Scientific interest: aspects of Applications of System Dynamics and Quantitative analysis of macroeconomic
Experience: He participated in many research projects about Private Economy from 2006 to 2011, mainly responsible for modelling work

Sun Qiming, China

Current position, grades: full professor of Management Science and Engineering at Economics and Management Department, Beijing University of Posts and Telecommunications. Since 2005, he is dean of Literature Law and Economics Department in BUPT
University studies: Bachelor degree in Political Economy (1983), named Doctoral tutor in Regional Economics (2003)
Scientific interest: different aspects of Balanced Development of Regional Economy and Management Research of Private Economy Contemporary in China

The buyback contract coordination for a logistics service supply chain

Wang Fen[1*], Chen Jianping[2]

[1] *School of management, Zhejiang Ocean University, Zhoushan, China*

[2] *School of Management, NingBo Institute of Technology, Zhejiang University, NingBo, China*

Received 1 January 2014, www.tsi.lv

Abstract

This article is about the coordination issue of the logistics service supply chain leading by the functional logistics provider (FLP). The service supply chain is consisted of the risk-neutral FLP and the loss-averse logistics integrator (LI), and the contract model of the wholesale price and buyback contract model are established. The study found that the wholesale price contract cannot coordinate the supply chain, but the introduction of the buyback contract can stimulate the LI to increase the order quantity of the logistics capacity, reaching the level of the centralized logistics service supply chain and finally it is verified through examples.

Keywords: logistics service supply chain, loss averse, buyback contract, coordination

1. Introduction

The logistics service supply chain (LSSC) is a new service supply chain, which is composed with logistics service integrators (LSI) and the functional logistics provider (FLP). LSI integrates all kind of FLPs' resources such as warehouse, transportation, distribution in order to supply integrated logistics services to customers (Choy et al. 2007) [1]. The logistics service supply chain is formed in order to adapt to the needs of the customers and the growth and development of the logistics enterprises. In essence, the LSSC is a service supply chain on the basis of capability collaboration (Lisa et al., 2004) [2], and its key operation problems is coordinating all of the enterprises in the chain. The supply chain contract is an important method to achieve the supply chain coordination [3].

Berglund (2000) researched the cooperation of logistics enterprises, he designed such buyback contract, i.e. when the third party logistics provider provides higher purchase price, the FLP buys back the unused service capacity with certain price discount to encourage the cooperation intent of the third logistics provider [4]. Liu (2008) considered the two-echelon supply chain structure with single period consisting of a FLP and a LI, established the LSI's cost model and the FLP's profits model with or without the capability collaboration restraint respectively, and also gave the capability coordination model under Stackelberg decision [5]. Gui et al. (2009) studied the two echelon logistics supply chain with one logistics service integrator and one functional logistics service provider, and developed the mathematical model of centralized coordination, Stackelberg game coordination and competitive aligned coordination based on the market characterized by a price sensitive random demand [6]. Liu (2010) studied the optimal revenue-sharing coefficient in three echelon logistics service supply chain in a stochastic demand environment [7]. Hu et al. (2011) studied the two echelon logistics supply chain with one logistics service integrator and one functional logistics service provider, and studied the coordination of buyback contract considering different quantity discount [8]. Liu et al. (2012) studied the quantity coordination of capability collaboration for multi-period-oriented two-echelon logistics service supply chain with Stackelberg decision-making [9]. Gui et al. (2012) studied the coordinating problem of logistics service supply chain under uncertain supply capacity, and proposed a payback contract to coordinate logistics service supply chain under deterministic and stochastic demand [10]. All these papers studied the quality coordination contract, quantity discount contract, buyback contract, revenue sharing contract in two or three echelon LSSC, but these models are based on the risk neutrality. They did not consider the behavioural issues in LSSC. Therefore, it is significant to study the coordination of LSSC considering the behavioural issues. This article takes the viewpoint of prospect theory rather than risk neutrality to describe the LSI's decision-making behaviour in a FLP-leading logistics service supply chain [11]. The article studies the wholesale price contract coordination and buyback contract coordination considering LSI as a loss –averse decision maker.

* Corresponding author. Tel: +86-135-6765-6834; fax: +86-0580-818-0411;
E-mail: worphan77@163.com

2. Decision Model of the Logistics Service Supply Chain

2.1. BASIC ASSUMPTIONS

This article takes into consideration of such situation in the logistics service market, i.e. certain FLP has absolute market advantage within its covered area coverage. When it joins in a certain logistics service supply chain as a FLP, the LI must cooperate with this powerful FLP out of the necessity of business development. Therefore, the logistics service supply chain leading by the functional logistics provider is formed [12].

Taking into consideration of a system model of LSSC with two stages and single cycle, this supply chain is consisted of two members, including the upstream FLP and downstream LI, and they can only have trade with single logistics capacity. The LI has the random logistics service needs. If the market demand is a non-negative and consecutive random variable, with the mean value of μ, the probability density function of $f(x)$ and cumulative distribution function of $F(x)$. F^{-} represents the inverse function of the cumulative distribution function of the random demand. All of the market demands of the LI are ordered from the FLP. According to the result of the market demand forecast, the logistics capacity of the LI from the professional logistics provider is q of the order quantity, the FLP transfers the price w with unit capacity to provide such logistics capacity for the LI according to the capacity order requirement of the FLP and uses this logistics capacity to provide the relevant logistics services when it executes the logistics solution of the logistics demander. The costs related to the logistics capacity of the FLP include two parts: one part is the cost caused at the logistics capacity investment, with the discounted present value of c_{sf} and the other part is the logistics capacity operation cost of c_{sv} when the LI provides logistics service. The operation cost of the LI is c_I, when the logistics capacity is excess, the unused logistics capacity of the LI causes no operation cost. At the end of the cycle, the FLP buys back the unused logistics capacity of the LI with the price of b. If there is out of stock, the overall stockout loss of the supply chain is g, the stockout loss of the integrator is g_I and the stockout loss of the provider is g_s. The π represents the profit, U represents utility, $E(\pi)$ represents the expected profit and $E(U)$ represents the expected utility. The subscript c represents the centralized supply chain, I represents the LI, s represents the FLP, sc represents the decentralized supply chain and * represents the optimum strategy of the merchant. If the unit logistics capacity price p of the LI is exogenously given, it assumes that $p > w > c_{sf} + c_{sv}$, $p > w + c_I$, $w \geq b \geq c_{sv}$ in order to guarantee the profit of the integrator and FLP.

2.2. THE DECISION UNDER THE INTEGRATED LOGISTICS SERVICE SUPPLY CHAIN

First, the logistics service supply chain should be considered as a centralized decision system. It is assumed that it is the risk neutral. The cost of the centralized logistics service supply chain is the sum of the integrator cost and function provider cost, that is $c = c_I + c_{sf} + c_{sv}$. If there is any out of stock, the stockout loss of the centralized supply chain is the sum of the stockout loss of integrator and stockout loss of function provider, that is $g = g_I + g_s$. The profit of the centralized logistics service supply chain is

$$\pi_c = \begin{cases} (p - c_I - c_{sv})x - c_{sf}q, x \leq q \\ (p - c_I - c_{sv} - c_{sf})q - g(x - q), x > q \end{cases} \tag{1}$$

The expected profit of the centralized logistics service supply chain is

$$E(\pi_c) = \int_0^q [(p - c_I - c_{sv})x - c_{sf}q] f(x) d_x + \int_q^{\infty} [(p - c_I - c_{sv} - c_{sf})q - g(x - q)] f(x) d_x \tag{2}$$

Doing the first and second order derivative of q for (2), due to $\frac{\partial^2 E(\pi_c)}{\partial q^2} < 0$, $E(\pi_c)$ is the concave function of q. Therefore, the existing optimum order quantity makes the centralized logistics service supply chain get the maximum expected profit with $\frac{\partial E(\pi_c)}{\partial q} = 0$, and the optimum order quantity q_c^* of the centralized logistics service supply chain meets

$$F(q_c^*) = \frac{p - c_I - c_{sv} - c_{sf}}{p - c_I - c_{sv} + g} \tag{3}$$

The decision of the centralized logistics service supply chain provides an ideal decision result, providing a benchmark for the design of the coordination contract of the logistics service supply chain.

2.3. THE DECISION UNDER THE DECENTRALIZED LOGISTICS SERVICE SUPPLY CHAIN

It is assumed that the decision maker has loss aversion; the initial wealth is w_0 (at the beginning of the cycle). If the profit or loss at the end of the cycle is higher or lower than the initial level, the loss aversion function of the decision maker is piecewise linear.

$$U(w)=\begin{cases} w-w_0, w\geq w_0 \\ \lambda(w-w_0), w<w_0 \end{cases} \tag{4}$$

$\lambda \geq 1$ defines the level of the loss aversion, w is the final wealth at the end of the cycle. If λ=1, the decision maker has the risk neutral. The higher the value of λ, the higher loss aversion level of the decision maker. For easy application, $w_0=0$ is in the literature [13].

2.4. THE LI AND FLP'S DECISION UNDER WHOLESALE PRICE CONTRACT

In the logistics service supply chain managed by the FLP, the FLP is predominant. According to the viewpoint of behavioural agency theory, the diversity can be employed to decrease the risk, so it is assumed that the FLP is risk neutral. Whereas the LI is in the bad situation and cannot diversify, so it is assumed that the LI is loss aversion. It is assumed that the FLP first offers the wholesale price for the LI; the LI determines the logistics capacity order quantity according to the wholesale price [14].

When the logistics capacity order quantity of the LI is q, the profit of the LI is

$$\pi_I(x,q,w)=\begin{cases} (p-c_I)x-wq, x\leq q \\ (p-c_I)q-wq-g_I(x-q), x>q \end{cases} \tag{5}$$

It is assumed that the corresponding market demand of the profit break-even point of the LI is q_I, $\pi_I(x,q,w)=0$, the profit break-even point of $q_I=\dfrac{w}{p-c_I}q$ and $q_I=\dfrac{p-c_I-w+g_I}{g_I}q$ can be got.

If $k_1=\dfrac{w}{p-c_I}$, $k_2=\dfrac{p-c_I-w+g_I}{g_I}$, when $x\in[0,k_1q)$ and $x\in(k_2q,\infty)$, $\pi_I<0$.

The expected profit function of the LI is

$$E[\pi_I(x,q,w)]=\int_0^q[(p-c_I)x-wq]f(x)dx+\int_q^\infty[(p-c_I-w)q-g_I(x-q)]f(x)dx \tag{6}$$

The expected utility loss of the LI is

$$L_I(q,w)=(\lambda_I-1)\int_0^{k_1q}[(p-c_I)x-wq]f(x)dx+(\lambda_I-1)\int_{k_2q}^\infty[(p-c_I-w)q-g_I(x-q)]f(x)dx \tag{7}$$

The expected utility of the LI is

$$E[U(\pi_I(x,q,w))]=E[\pi_I(x,q,w)]+L_I(q,w) \tag{8}$$

The decision goal of the LI is to maximize its expected utility in the context of given wholesale price by the function provider and doing the first and second derivative of q for (8).

$$\frac{\partial E[U(\pi_I(x,q,w))]}{\partial q}=(p-c_I-w+g_I)[\overline{F}(q)+(\lambda_I-1)\overline{F}(k_2q)]-w[F(q)+(\lambda_I-1)F(k_1q)] \tag{9}$$

$$\frac{\partial^2 E[U(\pi_I(x,q,w))]}{\partial q^2}=-(\lambda_I-1)\frac{w^2}{p-c_I}f(k_1q)-(p-c_I+g_I)f(q)-(\lambda_I-1)\frac{(p-c_I-w+g_I)^2}{g_I}f(k_2q)<0 \tag{10}$$

$E[U(\pi_I(x,q,w))]$ is the concave function of q, if $\dfrac{\partial E[U(\pi_I(x,q,w))]}{\partial q}=0$ in (9), the gained optimum logistics capacity order quantity q_I^* of the LI under the wholesale price contract meets

$$(p-c_I-w+g_I)[\overline{F}(q_I^*)+(\lambda_I-1)\overline{F}(k_2q_I^*)]-w[F(q_I^*)+(\lambda_I-1)F(k_1q_I^*)]=0 \tag{11}$$

Using the optimum order quantity q_I^* under the wholesale price contract of the LI to do the derivative for the stockout loss g_I can get

$$\frac{\partial q_I^*}{\partial g_I}=\frac{\partial^2 E[U(\pi_I(x,q,w))]/\partial q_I^*\partial g_I}{-\partial^2 E[U(\pi_I(x,q,w))]/\partial q_I^{*2}}>0 \tag{12}$$

Equation (12) shows that the optimum order quantity of the LI will increase along with the increase of the stockout loss under the wholesale price contract.

When the logistics capacity order quantity of the LI is q_I^*, the profit of the FLP is

$$\pi_s(w)=\begin{cases} wq_I^*-c_{sf}q_I^*-c_{sv}x, x\le q_I^* \\ (w+g_s-c_{sf}-c_{sv})q_I^*-g_sx, x>q_I^* \end{cases} \quad (13)$$

If the FLP is the risk neutral, the expected profit of the FLP is

$$E[\pi_s(w)]=\int_0^{q_I^*}[(w-c_{sf})q_I^*-c_{sv}x]f(x)dx + \int_{q_I^*}^{\infty}[(w+g_s-c_{sf}-c_{sv})q_I^*-g_sx]f(x)dx \quad (14)$$

The decision goal of the FLP is to offer the wholesale price for the LI to make its profit maximum. The order quantity q of the LI is the function of the wholesale price w, using the expected profit of the FLP $E[\pi_s(w)]$ to do the first derivative for w, can get

$$\frac{\partial E[\pi_s(w)]}{\partial w}=q_I^*+[w-c_{sf}+(g_s-c_{sv})\overline{F}(q_I^*)]\frac{\partial q_I^*}{\partial w}. \quad (15)$$

The optimum wholesale price w^* of the FLP meets (16).

$$q_I^*+[w^*-c_{sf}+(g_s-c_{sv})\overline{F}(q_I^*)]\frac{\partial q_I^*}{\partial w}=0 \quad (16)$$

Theorem 1. If the LI and FLP both have the optimum strategy, the order quantity of the LI will decrease along with the increase of the wholesale price.

It can be known from (16), the first item on the left side is positive number and the second item is positive number, so $\partial q_I^*(w)/\partial w<0$, the order quantity decreases along with the increase of the wholesale price.

If the order quantity of the loss-averse LI $q_I^*=q_c^*$, there must be

$$(p-c_I-w+g_I)[\overline{F}(q_c^*)+(\lambda_I-1)\overline{F}(k_2q_c^*)] -w[F(q_c^*)+(\lambda_I-1)F(k_1q_c^*)]=0 \quad (17)$$

$$A(g_I)=(p-c_I-w+g_I)[\overline{F}(q_c^*)+(\lambda_I-1)\overline{F}(k_2q_c^*)] -w[F(q_c^*)+(\lambda_I-1)F(k_1q_c^*)]$$

when $g_I\to 0$, $k_2q_c^*\to+\infty$, $\overline{F}(k_2q_c^*)\to 0$, if you want $A(g_I)<0$, then

$w<[(p-c_I)\overline{F}(q_c^*)]/[(\lambda_I-1)F(k_1q_c^*)+1]$, let

$w_0\in[c_{sf}+c_{sv},(p-c_I)\overline{F}(q_c^*)]/[(\lambda_I-1)F(k_1q_c^*)+1]$, $w=w_0$, then $\lim_{g_I\to 0}A(g_I)<0$.

When $g_I\to+\infty$, $A(g_I)\to+\infty$, then $\lim_{g_I\to+\infty}A(g_I)>0$, because the optimum order quantity of the LI increase along with the increase of the stockout loss, for w_0, g_I^0 must exist and $A(g_I^0)=0$. When $g_I\in(0,g_I^0)$, the logistics capacity order quantity of the LI $q_I^*<q_c^*$; when $g_I\in(g_I^0,+\infty)$, the logistics capacity order quantity of the LI $q_I^*>q_c^*$. Now, people pay much attention to $g_I\in(0,g_I^0)$, the wholesale price contract cannot coordinate LSSC, so consider to introduce the buyback contract to know whether it can coordinate the supply chain.

2.5. THE LI AND FLP'S DECISION UNDER BUYBACK CONTRACT

When the wholesale price contract cannot coordinate the LSSC, consider introducing the buyback contract (w_b,b), b represents the buyback price of the FLP and the coordination condition of the LSSC under the contract is researched.

When the FLP offers the contract (w_b,b), the logistics capacity order quantity of the LI is q_b and the profit of the LI is

$$\pi_I(x,q,w_b,b)=\begin{cases}(p-c_I-b)x-(w_b-b)q_b, x\le q_b \\ (p-c_I-w_b+g_I)q_b-g_Ix, x>q_b\end{cases}. \quad (18)$$

It is assumed that the corresponding market demand of the profit break-even point of the LI is q_I, $\pi_I(x,q,w_b,b)=0$, the profit beak-even point of $q_I=\frac{w_b-b}{p-c_I-b}q_b$ and $q_I=\frac{p-c_I-w_b+g_I}{g_I}q_b$ can be got. If $k_{1b}=\frac{w_b-b}{p-c_I-b}$, $k_{2b}=\frac{p-c_I-w_b+g_I}{g_I}$, when $x\in[0,k_{1b}q_b)$ and $x\in(k_{2b}q_b,\infty)$, $\pi_I<0$.

The expected profit function of the LI is

$$E[\pi_I(x,q_b,w_b,b)] = \int_0^{q_b}[([p-c_I-b)x-(w_b-b)q_b]f(x)dx + \int_{q_b}^{\infty}[(p-c_I-w_b+g_I)q_b-g_I x]f(x)dx \quad (19)$$

The expected utility loss of the LI is

$$L_I(q_b,w_b,b)=(\lambda_I-1)\int_0^{k_{1b}q_b}[(p-c_I-b)x-(w_b-b)q_b]f(x)dx + (\lambda_I-1)\int_{k_{2b}q_b}^{\infty}[(p-c_I-w_b+g_I)q_b-g_I x]f(x)dx \quad (20)$$

The expected utility of the LI under the buyback contract is

$$E[U(\pi_I(x,q_b,w_b,b))]=E[\pi_I(x,q_b,w_b,b)]+L_I(q_b,w_b,b) \quad (21)$$

The decision goal of the LI is to maximize its expected utility in the context of given wholesale price and buyback price by the FLP and doing the first and second derivative of q_b for (22)

$$\frac{\partial E[U(\pi_I(x,q_b,w_b,b))]}{\partial q_b}=(p-c_I-w_b+g_I)[\overline{F}(q_b)+(\lambda_I-1)\overline{F}(k_{2b}q_b)] - (w_b-b)[F(q_b)+(\lambda_I-1)F(k_{1b}q_b)] \quad (22)$$

$$\frac{\partial^2 E[U(\pi_I(x,q_b,w_b,b))]}{\partial q_b^{\,2}}=-(\lambda_I-1)\frac{(p-c_I-w_b+g_I)^2}{g_I}f(k_{2b}q_b) - (p-c_I+g_I-b)f(q_b)-(\lambda_I-1)\frac{(w_b-b)^2}{p-c_I}f(k_{1b}q_b) \quad (23)$$

It can be known from (23) that $\frac{\partial^2 E[U(\pi_I(x,q_b,w_b,b)]}{\partial q_b^{\,2}}<0$, $E[U(\pi_I(x,q_b,w_b,b))]$ is the concave function of q_b, if $\frac{\partial E[U(\pi_I(x,q_b,w_b,b))]}{\partial q_b}=0$ in (21), the gained optimum logistics capacity order quantity q_b^* of the LI under the buyback contract (w_b,b) meets (24) that is

$$(p-c_I-w_b+g_I)[\overline{F}(q_b^*)+(\lambda_I-1)\overline{F}(k_{2b}q_b^*)] - (w_b-b)[F(q_b^*)+(\lambda_I-1)F(k_{1b}q_b^*)]=0 \quad (24)$$

When the logistics capacity order quantity of the LI is q_b, the profit of the FLP is

$$\pi_s(w,b)=\begin{cases}(w_b-c_{sf}-b)q_b-(c_{sv}-b)x, x\le q_b\\(w_b+g_s-c_{sf}-c_{sv})q_b-g_s x, x>q_b\end{cases} \quad (25)$$

The expected profit of the FLP is

$$E[\pi_s(w_b,b)]=\int_0^{q_b}[(w_b-c_{sf}-b)q_b-(c_{sv}-b)x]f(x)dx + \int_{q_b}^{\infty}[(w_b+g_s-c_{sf}-c_{sv})q_b-g_s x]f(x)dx \quad (26)$$

Theorem 2. The buyback contract can coordinate the logistics service supply chain leading by the FLP.

Demonstration: in the buyback contract (w_b,b), put $q_b^*=q_c^*$ into (23) can get (26)

$$b(w_b)=\frac{(p-c_I-w_b+g_I)[\overline{F}(q_c^*)+(\lambda_I-1)\overline{F}(k_{2b}q_c^*)]}{F(q_c^*)+(\lambda_I-1)F(k_{1b}q_c^*)} - \frac{w_b[F(q_c^*)+(\lambda_I-1)F(k_{1b}q_c^*)]}{F(q_c^*)+(\lambda_I-1)F(k_{1b}q_c^*)} \quad (27)$$

Put $b(w_b)$ into (26), when $\frac{\partial E[\pi_s(w_b,b)]}{\partial w_b}=0$, you can get the value of w_b^* and b^*. It shows that the buyback contract can coordinate the logistics service supply chain.

According to the above-mentioned process, the analytic solution of the optimum wholesale price and buyback price can be got.

3. Numeric Analysis

It is assumed that the market demand is object to the uniform distribution $x\in U[0,5000]$, the retail price $p=14$, the stockout loss of the LI $g_I=2$, the operation cost of the integrator $c_I=1$, the stockout loss of the FLP $g_s=2$, the unit logistics capacity discount cost of the FLP $c_{sf}=3$, the unit logistics capacity operation cost $c_{sv}=1$, the loss aversion coefficient of the LI λ_I and the FLP is risk neutral. If the loss aversion coefficient of the LI λ_I is 1.5, 2, 2.5, 3, 4 respectively, it can be calculated that the optimum order quantity q_c^* of the centralized logistics service supply chain is 2,183 and the expected profit is 1,3905.4. You can get the optimum order

quantity and optimum wholesale price under the wholesale price contract from (11) and (16), please see Figure 1, from which you can know that the order quantity of the LI is less than the optimum order quantity of the centralized supply chain. Figure 1 verifies the theorem 1.When the LI and FLP both have the optimum strategy under the wholesale price contract, the order quantity of the LI decreases along with the increase of the wholesale price.

Under the buyback contract, the wholesale price and buyback price can be got through calculation, please see Figure 2. Along with the increase of the loss aversion coefficient, the wholesale price increases and the buyback price decreases.

The expected utility of the FLP under the wholesale price contract can be got through the calculation of (14), and the expected utility of the FLP under the buyback contract can be got through the calculation of (26), see Figure 3, which shows that under these two contracts, the expected utility of the provider increases along with the increase of the loss aversion coefficient of the LI; when the loss aversion coefficient is identical, the expected utility of the FLP under the wholesale price contract is smaller than that under the buyback price contract.

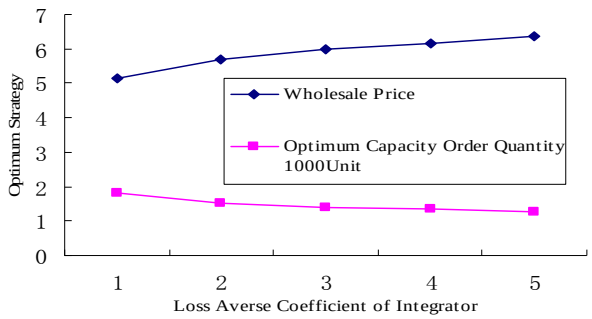


FIGURE 1. Optimum strategy under wholesale price contract

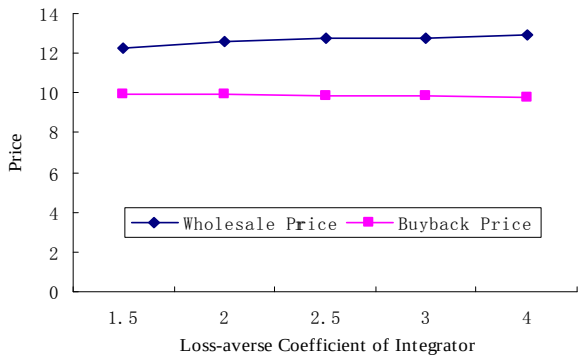


FIGURE 2. Loss-averse coefficient and price

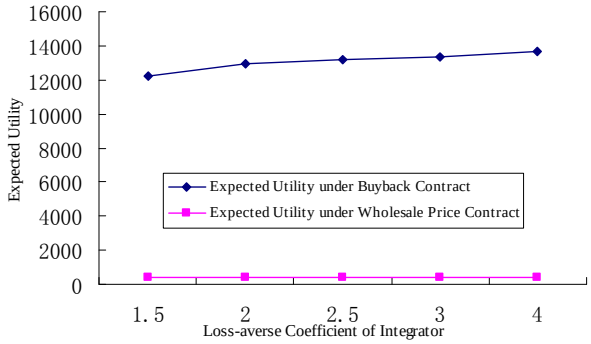


FIGURE 4. Loss-averse coefficient of LI and expected utilility of FPL

The total expected utility of the supply chain under the two contracts can be got through calculation, please see Figure 4, from which we can know that, 1) the total expected utility of the whole supply chain decreases along with the increase of the loss-averse coefficient under these two contracts, 2) when the loss-averse coefficient is identical, the total expected utility of the supply chain under the wholesale price contract is smaller than that under the buyback contract.

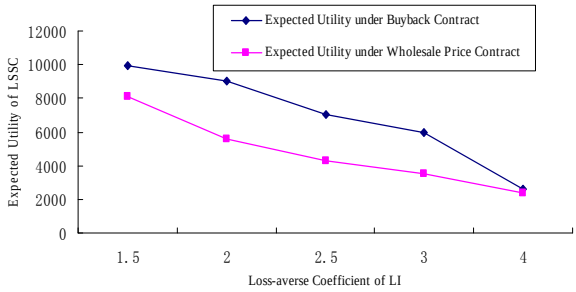


FIGURE 4. Loss-averse coefficient of LI and expected utilility

This example shows that the buyback contract can coordinate the logistics service supply chain and encourage the LI to order according to the optimum logistics capacity order quantity of the centralized logistics service supply chain and verifies the theorem 2.

4. Conclusions

This article researches the coordination issue of the buyback contract with two-layer logistics service supply chain and single stage managed by the FLP and the LI is considered to be the decision maker with loss aversion. The research finds that due to the loss–averse characteristic of the LI, its capacity order quantity is lower than the optimum capacity order quantity of the centralized supply chain under the wholesale price contract. The introduction of the buyback contract can encourage the LI to order according to the optimum capacity order quantity of the centralized supply chain to coordinate the logistics service supply chain. The FLP gets more utilities by using its predominant role. The two-layer logistics service supply chain coordination

leading by the LI taking into consideration of the behaviour factor, three-layer logistics service supply chain coordination taking into consideration of the behaviour factor, multiple stages coordination of the logistics service supply chain coordination taking into consideration of the behaviour factor and the logistics service supply chain coordination taking into consideration of the behaviour factor of the decision maker under the condition of information asymmetry can be further researched in future.

Acknowledgment

This work was financially supported by Soft Science Key Project in Zhejiang under grant No. 2012C25099，the National Natural Science Foundation of China under grant No. 71271191 and No. 70871111.

References

[1] Choy K. L., Chung-Lun Li, CKS Stuart, Henry L. (2007) *International J. Risk Assess. Manage* **, 7**(1), 19-43

[2] Ellram L. M., Tate W. L., Billington C. (2004) *J. Supply Chain Manage*, **40**(4), 17-32

[3] Hai S., Hu H. Z., Dai W. H., Huajuan Mao, Zhang Y. (2012) *J. Comput* , **7**(11), 2842-49

[4] Berglund M. (2000) *Strategic positioning of the emerging third party logistics providers Detoral dissertation*, LinkoPing Studies in Management and Economics

[5] WH Liu (2008) *J. Wuhan Univ. Technol.*, **30**(2), 149-53

[6] YM Gui, BG Gong, YM Chen (2009) *Comput Integ. Manuf. Syst*., **15**(12), 2412-16

[7] WH Liu (2010) *J. Southwest Jiaotoong Univ.*, **40**(5), 811-16

[8] CZ Hu, BG Gong, X Chen (2011) *J. Anhui Polytechnic Univ.*, **26**(4), 85-88

[9] W. Liu, X. Xu (2012) *Afr. J. Bus. Manag.*, **6**(11), 3920-9

[10] YM Gui, BG Gong, TL Zhang *J. Beijing Jiaotong Univ.Soc.*, **11**(2), 27-31, 2012

[11] XL Qu, ZX Hao (2011) *J. Comput*., **6**(11), 2454-2459

[12] WD Zhao, HF Wu, WH Dai, X Li, et al. (2011), *J.C omput*., **6**(7), 1469-76

[13] Charles X., Wang S., Webster (2012), *Decis.Sci.*, **38**(3), 361-389

[14] Hu H. Z., Dai W. H., Dai Y. H., Sun H. et al. (2012), *J. Softwar*., **7**(11), 2630-39

Authors

Wang Fen, 1977, born in Hubei, China

Current position, grades: a lecturer in the school of management of Zhejiang Ocean University.
University studies: is a doctoral candidate in the business administration of Zhejiang Gongshang University
Scientific interest: logistics service supply chain and port logistics

JianPing Chen, 1977, in Jiangxi, China

Current position, grades: a lecturer in NingBo Institute of Technology Zhejiang University.
University studies: doctor of management
Scientific interest: is technology innovation and service innovation

Research on resource-constrained project scheduling method based on heuristic priority rules

Bing Wang[1], Tinggui Chen[2 *], Guanglan Zhou[3], Chonghuan Xu[4]

[1] *Hangzhou College of Commerce, Zhejiang Gongshang University, Hangzhou, P. R. China*

[2] *College of Computer Science & Information Engineering, Zhejiang Gongshang University, Hangzhou, P. R. China*

[3] *Center for Studies of Modern Business, Zhejiang Gongshang University, Hangzhou, P. R. China*

[4] *School of Business Administration, Zhejiang Gongshang University, Hangzhou, P. R. China*

Received 1 January 2014, www.tsi.lv

Abstract

The traditional project scheduling problem only considers the logical constraints. However, there is a need to consider the resources such as labour, capital and other constraints. Resource constrained project scheduling problem (RCPSP) integrates logic and resource constraints, which are closer to the actual scheduling. And most of these problems belong to NP-hard problem, which have certain difficulty in solving process. And therefore, there is a very important significance to study in the RCPSP, especially in the theory and application. Among the many the RCPSP models, resource constrained project scheduling problem is the most fundamental, but also the most basic model. Most RCPSP research has focused on a single project scheduling problem. The main content of this article describes the single resource constrained project scheduling problem and establishes the according model. Then, it studies how to use heuristic priority rules for solving the single project scheduling problem. In addition, this paper finally simulates a single project scheduling as an example of mold production and solves it by using the heuristic algorithm based on priority rules in order to verify the effectiveness of the algorithm. It combines with the different schedule generation schemes and priority rules as well as compares the different solution results. The final outcome indicates that the combination of different priority rules and schedule generation schemes would influence the single project scheduling results.

Keywords: resource constraints, project scheduling, heuristic priority rules

1. Introduction

Project scheduling is one of the most important procedures in project management. How to deal with the various tasks in the project schedule is crucial for the project management. Generally speaking, the project scheduling is based on the scheduling objective, based on certain human, material and financial resources, based on the sequence of execution of each task. Through the reasonable arrangement, a reasonable schedule scheme is eventually emerging.

Usually, the project scheduling tools include Gantt chart and network plan technology. The earliest one is the Gantt Chart [1]. Gantt chart indicates the execution order of the tasks and the time parameters, so that the project members can intuitively understand each task starting time and the order of execution, which can effectively evaluate the whole project scheduling process. The network planning technology [2] is the emerging technology developed in the mid 50's twentieth Century, then is quickly popular into the project scheduling field. With the development of network technology, it produces two core technologies: the critical path method, CPM and plan evaluation and review technique, PERT [3]. The network planning techniques could demonstrate the entire process of project scheduling, which could be used to describe the constraints between each task. Its structure is illustrated in figure 1. Other methods, such as critical path method, plan evaluation and review technique have some serious limitations in the project task scheduling. It assumes that the constraint does not exist in the task allocation process. Accordingly, this article adopts a heuristic method to solve the resource constrained project scheduling problem. The article is organized as follows: the second section discusses the single project scheduling problem and the corresponding mathematical model; the third section elaborates the heuristic algorithm for solving project scheduling problem based on priority rules; the fourth section verifies the proposed algorithm by introducing an example of the single project scheduling in mold production; the last one is the summary and analysis and prospects for future research.

* Corresponding author. Tel: +86-15268134776; fax: +86-0571-28004299;
E-mail: ctgsimon@gmail.com

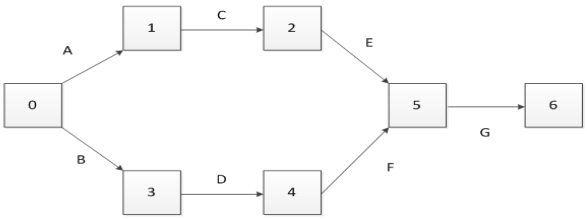


FIGURE 1. The network structure diagram.

2. An Analysis and Model Establishment of Project Scheduling Problem

Resource constrained single project scheduling problem (RCSPSP) [4] refers to the RCPSP problem which is for single project scheduling. The model is only applied to a project scheduling, logical and resource constraints between each task. It only considers renewable resources for each task specific constraints.

In a single project scheduling, the project includes J tasks. Structure for the project consists of a single code network chart (AON), where the nodes in the graph represent the task of the project, and directed arrows represent the constraint relationship between tasks. The first node and the last node are task 0 and task map in J respectively, which denotes the start time and end time for project. Moreover, execution time is 0 and there is no need to allocate resources. All tasks in the project have their corresponding duration d_j, start time ST_j and end time FT_j.

In the single project scheduling, there are constraints between each task. These constraints are usually referred to logical constraints and resource constraints. Logical constraints are described in specific performance for the precedence relation constraint among tasks. Each task must wait until its former task completed, i.e. the following up task is constraint by the former task. Secondly, as the unique to RCPSP resource constraints, they are usually referred that in the task execution process, each task needs to allocate different resources, and the number of resources is limited. The resource constraints are expressed as follows: in a K renewable resources in single project scheduling, the total resources of k(k=1,2,…,K) is R_k. The task j is required the renewable resources amount r_{jk} in the process of implementation. Then the constraint relationship is $\sum_{A_j \in I_t} r_{jk} \le R_k, \forall k, t$.

In resource constrained project scheduling problem, there is another important condition, besides the time parameters of the task constraint, which is objective function, the project optimization goal. At present, the objective function of time scope and resource scope are very popular in the single project scheduling. For the objective function of time scope, there are the minimum duration minFT_J, minimum project delay minmax{0,FT_j-d_j}, while for the objective function of resource scope, there is most common one minimum resource total cost $\min \sum_k C_k(R_k)$.

TABLE 1. The symbol lists of single project scheduling model

Symbol	Explanation
j	task index, j=1,2,…,J (J denote the total number of tasks)
t	time index, t=0,1,2,…,T (T denotes the upper limit of project makespan)
k	resource index, j=1,2,…,K (K denotes the total number of renewal resources required by a project)
d_j	the duration of task j
P_j	a immediate predecessor set of task j
S_j	a immediate successor set of task j
ST_j	the start time of task j
FT_j	the finished time of task j
R	the renewal resource set required by a project
R_k	the amount of resource k
r_{jk}	the amount of resource k take j requires
I_t	the execution task set at time t

According to the resource constrained project scheduling problem description, usually in a single project scheduling, the most common objective function is the minimum total duration. After meeting the logical constraints and resource constraints, there are reasonable arrangements for the task scheduling order, so that it produces the shortest total duration. The single project scheduling model in this paper proposes minimum total duration as its objective function. Accordingly, there are hypotheses below.

(1) For the precedence constraints, it only considers immediately executed task after the former task completed.

(2) The objective function is minimizing the project duration.

(3) Each project could not be interrupted, non suspension.

(4) The renewable resource constraints are concerned only.

The single project scheduling mathematical model of symbol is list in table 1. The mathematical model is established as follow.

$$\text{Min } FT_j \tag{1}$$

$$ST_j \ge FT_h, \forall h \in P_j, \forall j \tag{2}$$

$$\sum_{A_j \in I_t} r_{jk} \le R_k, \forall k, t \tag{3}$$

$$ST_j \ge 0, \forall j, \tag{4}$$

where formula (1) is the minimum total duration objective function; formula (2) is the precedence

relationship constraints, formula (3) is the resource constraint; formula (4) is the non negative constraint.

When solving the resource constraint of single project scheduling problem model, it is essential to find a suitable schedule for each task to determine a reasonable start time to meet the constraint conditions. The project schedule can be represented as an array, namely $S=(s_1, s_2,\ldots, s_J)$. Figure 2 shows a single project scheduling example of J=9. The project has a renewable resource only, i.e. K=1, the total amount of resources R=4. Each task's duration and resource requirements are remarked in the AON diagram. Task 1 and task 9 are the start task and the finish task respectively, which do not take up any resources.

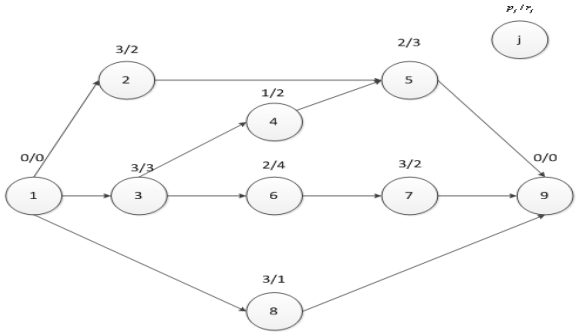


FIGURE 2. The single project scheduling paradigm.

3. Solution to the Project Scheduling Problem Based on Priority Rules

A heuristic algorithm is proposed by Kelly. Project scheduling is concerned with single-item or small batch production where scarce resources have to be allocated to dependent activities over time. More specifically, the exact and heuristic algorithms for the single-mode and the multi-mode case are reviewed, for the time–cost trade-off problem, for problems with minimum and maximum time lags, for problems with other objectives than make span minimization [5]. The well-known resource-constrained project scheduling problem (RCPSP) summarizes and categorizes a large number of heuristics that have recently been proposed in the literature. Most of these heuristics are then evaluated in a computational study and compared on the basis of our standardized experimental design. Therefore, the researchers have put forward the different solutions based on heuristic algorithm [6].

A new heuristic called self-adapting genetic algorithm employs the well-known activity list representation and considers two different decoding procedures. An additional gene in the representation determines which of the two decoding procedures is actually used to compute a schedule for an individual. This allows the genetic algorithm to adapt itself to the problem instance actually solved. Computational experiments show that the proposed heuristic is among the best ones currently available for the RCPSP [7]. Because this algorithm is relatively simple, it is quite popular. With the RCPSP expansion and model complexity, the efficiency of a heuristic algorithm for single channel is greatly reduced.

The project scheduling problem involves the scheduling of project activities subject to precedence and/or resource constraints. Despite all these efforts, numerous reports reveal that many projects escalate in time and budget and that many project scheduling procedures have not yet found their way to practical use. Herroelen [8] provides a generic hierarchical project planning and control framework that serves to position the various project planning procedures and discuss important research opportunities, the exploration of which may help to close the theory-practice gap [8]. The heuristic algorithm with various combinations of schedule generation mechanism and priority rule is called multi channel algorithm. The following ones are the common multi channel heuristic algorithms.

(1) Multi priority rule method. The method selects a schedule generation mechanism, and pursues the scheduling through the different priority rules.

(2) Forward and reverse backtracking method. This algorithm selects a schedule generation mechanism. After the repeated forward calculation and reverse calculation, it produces a series of project scheduling.

(3) Sampling algorithm. Sampling algorithm usually use only one schedule generation scheme and only one priority rule. However, in real situation, it determines the task scheduling via the probability of task priority value calculation, not according to the task priority value.

(4) Adaptive heuristic algorithm. The algorithm has high elasticity. It firstly selects the combination of schedule generation mechanism and priority rules according to the specific task quantity, constraint condition, and therefore it is greatly flexible.

Priority rule mechanism used in this paper is the generation mechanism of parallel and serial schedule based on the serial SGS with task as the phase variables, composed of J phases. Each phase of the g(g=1,2,…,J) corresponds to an incomplete plan task set PS_g and a feasible task set D_g, where the incomplete task set PS_g contains all the tasks which have arranged the start time, while the feasible task set D_g contains all the tasks which have not arranged the start time, but have already arranged all the preceding task, i.e. the task precedence task have been included in the incomplete task plan. At each new stage g, serial SGS would choose maximum priority coefficient task j* from D_g into the current PS_g according to the priority rules in order to meet the precedence relations and resource constraints, as well as assign the specified start time, ST_{j^*} for task j* and allocate resources. It repeats the scheduling, gradually extending PS_g, until the completion of the whole project schedule. Set all the tasks carried out in the time t as I_t, where $R_k(t) = R_k - \sum_{j\in I_t} r_{jk}$ is the remaining amount of the k resource at time t, and $v(j)$ is the priority

coefficient for task j in D_g. Consequently, the serial scheduling generation mechanism of the process is illustrated in Figure 3.

Initialization: g=1, FT_1=0,PS1={1};
While g<J do phase n
Begin
Calculate D_g, $R_k(t), k = 1,2,\ldots,K, t = FT_g$;

$j^* = \min_{j \in D_g}\{ j \,|\, v(j) = \max_{i \in D_g}\{v(i)\}\}$; //select the maximum priority coefficient task j* from Dg //

$ES_{j*} = \max\{FT_h \,|\, h \in P_j\}$; //Calculate the earliest start time of task j*//

$FT_{j^*} = \min\{t \,|\, t \geq ES_{j^*}, r_{kj} \leq R_k(\tau), k = 1,2,\ldots,K, \tau = t, t+1,\ldots,t+P_{j^*}\} + d_{j^*}$;//Calculate the finish time of task j*//

$PS_{g+1} = PS_g \cup \{j^*\}$;

g=g+1;
END;

FIGURE 3. The serial scheduling generation mechanism

Parallel scheduling generation mechanism adopts the time as phase variable, contains a maximum of J phases. Each phase g(g=1,2,…,J) corresponds to a scheduling time t_g. Set C_g as the completed task set at time t_g, and A_g as the proceeding task set at time t_g and D_g as the feasible task set at time t_g, where D_g contains all the tasks meeting the precedence relations and resource constraints which can be started at time t_g. At each stage, there are the following two steps for parallel schedule generation mechanism. (1) Determine the present time t_g. Exclude all the tasks of finish time equating to time t_g from A_g, then add them to the completed task set C_g, and update the feasible task set D_g. (2) In accordance with the principle of priority tasks, select the greater priority coefficient task j* from D_g, start the task from the current phase t_g, then shift the task j* from the feasible task set D_g to the proceeding task set A_g. Repeat step (2) until the feasible task set D_g is null, then move on to the next phase. Finally, when all the tasks have been belonged to the completed task set C_g or the proceeding task set A_g, the scheduling mechanism finishes. Set $R_k(t_g) = R_k - \sum_{j \in A_t} r_{jk}$ as the remaining amount of k resource at time t, where $v(j)$ is the priority coefficient D_g of task j. The parallel scheduling generation mechanism of the process is illustrated in Figure 4.

Initialization: g=1, t_g=0, D_g={1}, A_g=C_g=∅, $R_k(t_g) = R_k, k = 1,2,\ldots,K$;

While $|A_g \cup C_g| < J$ phase g

Begin

(1) $t_g = \min\{ST_j + d_j \,|\, j \in A_{n-1}\}$;

$A_n = A_{n-1} \setminus \{ j \,|\, j \in A_{n-1}, ST_j + d_j = t_g\}$;

$C_n = C_{n-1} \cup \{ j \,|\, j \in A_{n-1}, ST_j + d_j = t_g\}$;

Calculate $R_k(t_g)$ (k=1,2,…,K) & D_g;

(2) $j^* = \min_{j \in D_g}\{ j \,|\, v(j) = \max_{i \in D_g}\{v(i)\}\}$;

$ST_{j^*} = t_g$;

$A_g = A_g \cup \{j*\}$;

Calculate $R_k(t_g)$ (k=1,2,…,K) & D_g;

If $D_g \neq \varnothing$ Then GOTO Step(2) ELSE n=n+1;

END

FIGURE 4. The parallel scheduling generation mechanism

In addition, due to the serial and parallel scheduling process described above, there is a process in the selection of maximum priority coefficient task j^* from the feasible task set D_g. The value of the priority coefficient here is determined by the priority rules, and therefore there is a need for analysis on various priority rules. Specifically, priority rules refer to the expected generation mechanism that according to certain rules, each task priority coefficient is assigned in a feasible task set, which determines the task start time and the order of execution. There are four priority rules network-based rule, NBR, critical path-based rule, CPBR, resource-based rule, RBR, composite rule, CR. By using the iterative forward-backward scheduling technique, its application in real projects and comparison with other scheduling schemes confirmed that the proposed algorithm is capable to generate effective schedules for multiple projects with limited renewable resources [9].

4. Case Study

In order to verify the validity of the algorithm, the concrete examples of the resource constrained project scheduling problem in this section are listed. The heuristic algorithm based on priority rules provides solution and the comparison of serial schedule generation scheme and parallel schedule generation scheme, and the effects of different priority rules for different feasible solutions.

This scheduling model assumes that a single project scheduling has 7 tasks, the scheduling process only uses 1 renewable resource, and the renewable resources amount is 5. According to the precedence task constraint relation, the established network diagram is illustrated in figure 5. The task execution time *d* and resource usage *r* are listed in Table 2. Without taking into account the various task resource constraints, the earliest start time and latest start time are listed in Table 3. Under the condition of no resource constraints, the earliest finish time of completed project is for 28 days.

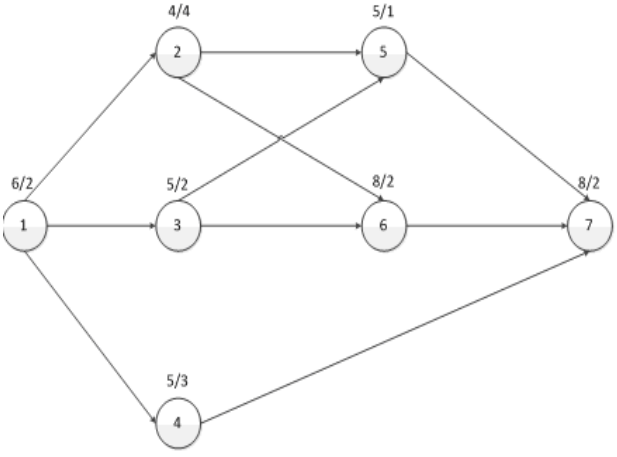


FIGURE 5. The case of single project scheduling

TABLE 2. Each task execution time and resource usage

Task *j*	Earliest start time *ES*	Latest start time *FS*
1	6	2
2	4	2
3	5	2
4	5	3
5	5	5
6	8	2
7	8	2

A feasible schedule serial schedule generation scheme with arbitrary priority rules can be obtained in this project. Here, firstly determine the shortest duration for the task priority rules. Then the algorithm process is described as follows.

The first phase: initialize the feasible task set D_1, D_1={1}. The task 1 is scheduling preferentially, start time is 0, the execution time is 6 days, the allocation of resources is 2.

TABLE 3. Each task earliest start time and latest start time

Task *j*	Execution time *d*	Resource usage *r*
1	0	0
2	6	7
3	6	6
4	6	14
5	11	11
6	11	14
7	19	19

The second phase: Task 2, 3, 4 consist of D_2, namely D_2={2,3,4}. According to the priority rule of the shortest duration, it schedules task 2. Set task 2 start time is 6, the execution time is 4 days, the allocation of resources is 2.

The third phase: D3={3,4}, the duration of task 3 and task 4 are both equal to 5. There is a need to use supplementary rules, namely the minimum number of priority scheduling rules. Therefore, the task 3 is scheduling preferentially. Set the start time is 10, the execution time is 5 days, the allocation of resources is 2.

The fourth phase: D_4={4,5,6}. According to the supplemental priority rule and minimum time priority rule, the task 4 is scheduling preferentially. Because the required resources of task 4 and task 3 both are 5, it can schedule the resource allocation simultaneously. Set start time is 10, the execution time is 5 days, the allocation of resources is 3.

The fifth phase: D_5={5,6}. According to the priority rules, the task 5 is scheduling preferentially. Set the start time is 15, the execution time is 5 days, the allocation of resources is 5.

The sixth phase: D_6={6}. Because the task 5 has exhausted 5 resources, the task 6 cannot share the resources with task 6. Set the start time is 20, the allocation of resources is 1.

The seventh phase: D_7={7}. After task 6 completing, task 7 is scheduling. Set the start time is 28, the allocation of resources is 1.

Through the above seven phases of scheduling, a feasible schedule set D_j is null. The scheduling finishes. A feasible planning project produces (represented by the task list form): {1,2,3,4,5,6,7}. The progress of the whole project arrangement is illustrated in figure 6.

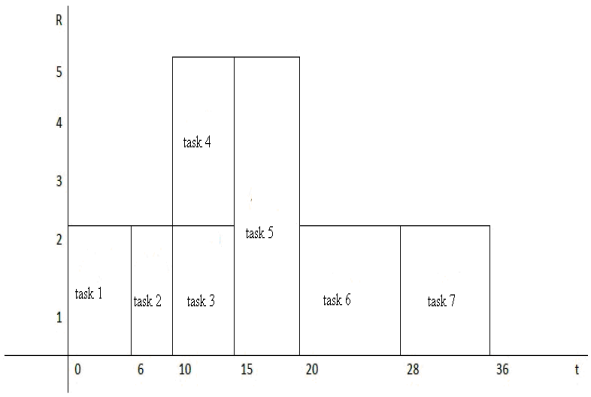


FIGURE 6. The whole project scheduling arrangement.

Secondly, it adopts the combination of the serial schedule generation scheme and the minimum time priority rule for the solution. The total time delay refers to the each task differences of the earliest start time and latest start or earliest finish time and latest finish time in project scheduling. According to the time parameter table of the project schedule in the former section, the critical path under the condition of non resource constraints is 1-3-5-7. The algorithm process is described below.

The first phase: D_1={1}. The task 1 is scheduling preferentially. The start time is 0, the execution time is 6, the allocation of resources is 2.

The second phase: D_2={2,3,4}. Based on priority rules, the task 3 is scheduling preferentially. The start time is 6, the execution time is 5, the allocation of resources is 2.

The third phase: D_3={2,4}. Based on priority rules, the task 2 is scheduling preferentially. The start time is 11, the execution time is 4, the allocation of resources is 2.

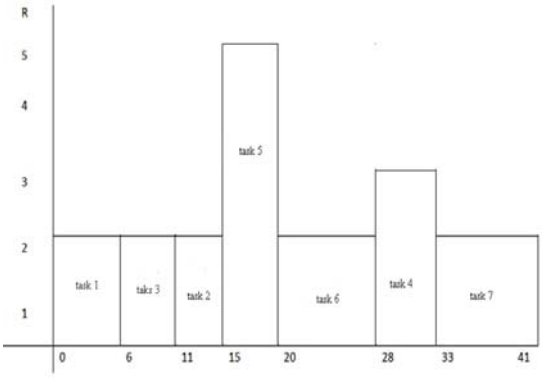
The fourth phase: D_4={4,5,6}. Based on priority rules, the task 5 is scheduling preferentially. The start time is 15, the execution time is 5, the allocation of resources is 5.

The fifth phase: D_5={4,6}. Based on priority rules, the task 6 is scheduling preferentially. The start time is 20, the execution time is 8, the allocation of resources is 2.

The sixth phase: D_6={4,7}. Based on priority rules, the task 4 is scheduling preferentially. The start time is 28, the execution time is 5, the allocation of resources is 3.

The seventh phase: D_7={7}. Based on priority rules, the task 7 is scheduling preferentially. The start time is 33, the execution time is 8, the allocation of resources is 2.

Through the calculation above, we can draw the conclusion that the feasible scheduling using the minimum total time difference priority rule is {1,3,2,5,6,4,7}.The exhaustive arrangement is illustrated in figure 7.



FIGURFE 7. The whole project scheduling arrangement.

The results show that under the same schedule generation mechanism, the different feasible schedules are calculated through priority rule. And thus, it will have a direct impact on the quality of project scheduling.

(2) The parallel schedule generation scheme for the solution

Here, another generation mechanism is used to solve the above examples, namely the use of parallel schedule generation scheme, and the selection of the most follow-up task as the priority rule. Let C_g the completed task set, A_g proceeding set at the time t_g, D_g the feasible task set at time t_g. The process is described as follows.

The first phase: D_1={1}. Allocate task 1. The current time is t_1=0, task 1 the start time is s_1= t_1=0, the resource allocation is 2, execute task set A_1={1}.

The second phase: the current time is t_2=6, D_2={2,3,4}. The subsequent task number of task 3 and task 2 is same. According to supplemental priority rule, task 2 is scheduling preferentially. The start time of task 2 is s_2=t_2=6, resource allocation is 2. After refreshing the remaining supply renewable resources, task 4 cannot continue to supply. So calculate D_2={3}. The task 3 is scheduling preferentially. Set the start time is s_3=t_2=6, the resource allocation is 2. Repeat calculation D_2=∅, enter the third phase.

The third phase: the calculation of the current time is t_3=11, D_3={4,5,6}, A_3=∅, C_3={1,2,3}. Task 4 is scheduling preferentially. Set the start time is s_4=t_3=11. Then A_3={4}, refresh remaining amount of renewable resources, task 5 cannot supply resources. And therefore, D_3={6}, task 6 is scheduling. Set the start time is s_6=t_3=11, then A_3={4,6}. Repeat calculation, D_3=∅, enter the next phase.

The fourth phase: the calculation of the current time is t_4=19, D_4={5}. The task 5 is scheduling preferentially. Set the start time is s_5=t_4=19, A_4={5}, the resource allocation is 5. Repeat calculation, D_4=∅, enter the next phase.

The fifth phase: the calculation of the current time is t_5=24, D_5={7}. The task 7 is scheduling preferentially. Set the start time is s_7=t_5=24, the resource allocation is 2. Then, finish the scheduling.

Through the calculation, we can draw a conclusion that the feasible scheduling using parallel schedule generation scheme and most subsequent tasks is {1,2,3,4,6,5,7}. The specific scheduling arrangement is illustrated in figure 8.

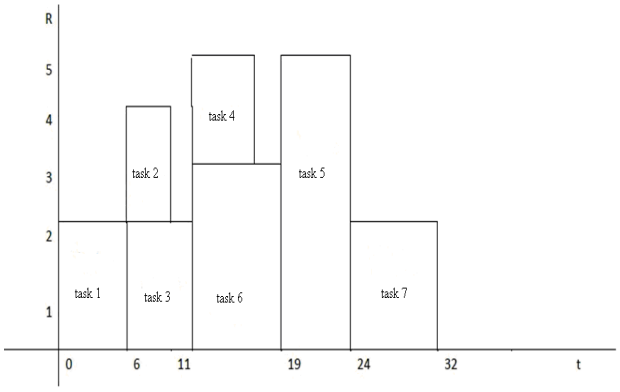


FIGURE 8. The whole project scheduling arrangement.

The examples model above verifies the feasible scheduling priority rule based heuristic algorithm which can obtain the resource constrained project scheduling. And the task scheduling would be different by using task serial and parallel generation mechanism respectively. Its practical application should be based on specific issues to choose the schedule generation scheme. Finally, the experiment also proves that the different priority rules will have an impact on the results under the same schedule generation mechanism.

5. Conclusions

Firstly, the traditional project scheduling problem background and related problem solving tools are introduced. Then, the single project scheduling problem and resource constraints are described. After that, a mathematical model is established. Based on the heuristic algorithm research on priority rules, two aspects from the schedule generation mechanism and priority rules are studied. The serial schedule generation scheme and parallel schedule generation scheme algorithm are introduced. There is an analysis of the characteristics of various priority rules. Finally, an example of mold production for single project scheduling is demonstrated. The example of the priority rule heuristic algorithm is presented. During the solution process, there are comparison and analysis of the different effects on the results of serial schedule generation scheme and parallel schedule generation mechanism. And there is another analysis on results via the combination of different schedule generation mechanism and different priority rule. After all, the heuristic algorithm based on priority rules indicates the validity single project scheduling problem under the resource constraints.

The flaw of the article is the selection of the heuristic algorithm based on priority rules for the solution to the single project scheduling problem. Although the use of the algorithm obtains the scheduling solution, for some complex project scheduling problem, the algorithm could not obtain the optimal solution, only a series of feasible solutions. In order to get the optimal solution, these feasible solutions are required to continue some subsequent operations. Therefore, the further studies are needed upon the intelligent algorithm from the aspect of algorithm, which could be applied to some complex project scheduling problem.

Acknowledgments

This research is supported by National Natural Science Foundation of China (Grant No. 71001088), Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20103326110001 and 20103326120001), Humanity and Sociology Foundation of Ministry of Education of China (Grant No. 11YJC630019), Zhejiang Provincial Natural Science Foundation of China (No. Y7100673 and LQ12G01007), the Scientific Research Fund of Zhejiang Province of China (Grand No. 2011C33G2050035), as well as Innovative Group of e-Business Technology of Zhejiang Province (2010R50041), Foundation of Department of Education of Zhejiang Province(Y201224109 and Y201329545), Key Laboratory of Electronic Business and Logistics Information Technology of Zhejiang Province (2011E10005).

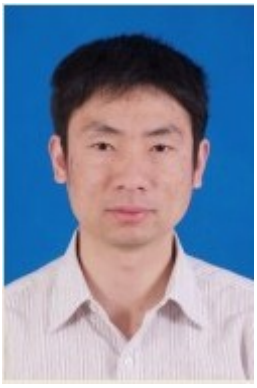
References

[1] Liu Q. (2007) *Project optimization scheduling theory and method.* Beijing: Mechanics and Industrial Press, chapter 5
[2] Shou Y. (2010) *The model and method for resource constrained multi project scheduling problem.* Hangzhou: Zhejiang: Zhejiang University Press, chapter 3
[3] Ying Y. (2010) *The research of uncertain resource constrained project scheduling problem.* Zhejiang: Zhejiang University Press, chapter 2
[4] Liu S., Wang M., Tang J. (2001) *Control and Decision*, **16,** 647-51
[5] Brucker P., Drexl A., Mohring M. (1999) *European Journal of Operational Research*, **112**(1), 995603-41
[6] Kolisch R., Hartmann S. (2006) *European Journal of Operational Research*, **174**(1), 061723-37
[7] Hartmann S. A. (2002) *Naval Research Logistics*, **49**(5), 023433-48
[8] Herroelen W. (2006) *Production and Operation Management*, **14**(4), 063412-32
[9] Shou Y. (2004) *Journal of Zhejiang University*, **38**(8), 1095-99

Authors

Bing Wang, born 06.07.1981, Hangzhou, P. R. China

Current position, grades: lecturer, Ph. D. Zhejiang Gongshang University
University studies: Computer science and technology
Scientific interest: Computer modelling
Publications: 5
Experience: Dr. Bing Wang is a lecturer of Hangzhou College of Commerce at Zhejiang Gongshang university. He was born in 1981. He got his Ph.D. degree in Enterprise Management from Zhejiang Gongshang University. His current research interest focuses on enterprise information system, electronic commerce, mobile commerce and context-awareness. His email is iceking0706@mail.zjgsu.edu.cn.



Tinggui Chen, born 06.04.1979, Wuhan, P. R. China

Current position, grades: associate professor, Ph. D. Zhejiang Gongshang University
University studies: Operations research
Scientific interest: Operations research
Publications: 30
Experience: Professor Tinggui Chen is an assistant professor of Information System, College of Computer and Information Engineering at Zhejiang Gongshang university. He was born in 1979. He got his Ph.D. degree in Industrial engineering from Huazhong University of Science and Technology. His current research interest focuses on management decision theory and decision support systems, swarm intelligence and complexity science. He has published over 30 publications in academic journals and conference proceedings. His email is ctgsimon@gmail.com.

Guanglan Zhou, 17.03.1983, Hangzhou, P. R. China

Current position, grades: lecturer, Zhejiang Gongshang University
University studies: Operations research
Scientific interest: Operations research
Publications: 10
Experience: Dr. Guanglan Zhou is a lecture of Information System, College of Computer and Information Engineering at Zhejiang Gongshang university. He was born in 1983. He received his Master of Science degree in 2007, conferred by the University of Liverpool, UK. Additionally, he was employed as an Academic secretary, by National Science Foundation of China, NSFC from 2010 to 2011. His research interests include data mining, electronic commerce and supply chain management. He has published over 10 publications in academic journals and conference proceedings. His email is guanglanzhou@mail.zjgsu.edu.cn

Chonghuan Xu, 27.04.1983, Hangzhou, P. R. China

Current position, grades: lecturer, Zhejiang Gongshang University
University studies: Operations research
Scientific interest: Operations research
Publications: 15
Experience: Dr. Chonghuan Xu is a lecture of College of Business Administration Zhejiang Gongshang University. He was born in 1983. He received his B.S. and M.S. degrees in Computer and Information Engineering from Zhejiang Gongshang University, Hangzhou. His research interests include electronic commerce, data mining. He has published over 15 publications in academic journals and conference proceedings. His email is talentxch@gmail.com.

Optimization research of material inventory management based on genetic algorithm

Xie Yuanmin[1*]

[1] *The Key Laboratory for Metallurgical Equipment and Control of the Ministry of Education, Wuhan University of Science and Technology, Wuhan, 430081, China*

Received 1 January 2014, www.tsi.lv

Abstract

Inventory management is an indispensable part in supply chain management. On the one hand the important position it has in the enterprise competitiveness. It has a direct impact on the cost of the products, respond speed to the market, delivery date and other indexes; On the other hand, the inventory system is a dynamic system, which involves huge inventory and wide area. The paper first under the influence of the uncertain factors does certain analysis to the raw material inventory issues, establishes minimum model of raw materials inventory cost in the iron and steel enterprises. Among them, the raw materials inventory cost includes fixed cost, procurement cost and storage cost. By making each purchase amount and raw materials inventory as the decision-making variables, it adopts the corrected genetic algorithm and uses MATLAB to get the optimal solution and get the most optimized raw materials procurement and inventory. The optimization model of raw materials inventory control which presents in this paper mainly applies to iron and steel enterprises of continuous production process, and can also be extended to other types of the enterprises in raw materials inventory control.

Keywords: material inventory, iron and steel industry, optimization, genetic algorithm, MATLAB

1. Introduction

For the characteristics of continuous production of iron and steel enterprises, the problem of material inventory is worthy of studying: on the one hand, raw material inventory takes up a lot of money, impacting on the speed of capital turnover; on the other hand, the uncertainty exists in the procurement, such as not timely supply, a shortage of raw materials, etc., and raw material purchasing cost accounts for 60% to 70% of the total cost of the iron and steel production enterprise, it is crucial to determine reasonable the material inventory and purchasing quantity in order to ensure the normal and continuous production of enterprise.

Iron and steel industry is the foundation of the industrial economy, undertaking the important task of providing raw materials to other industry sectors, which has made a significant contribution to promoting the development and progress of the industrial economy [1]. At present, the type of iron and steel enterprise is converting from the extensive type to technology-intensive type with the intense competition of global steel market and the influence of limited resources, in order to improve the competitiveness of products and grab a bigger share of the market, steel enterprises must constantly compress production costs, improve product quality, improve the level of customer service, speed up the turnover of funds and make the whole supply chain coordination. However, raw material inventory and purchasing problem of iron and steel enterprise has some characteristics, such as kinds of varieties, large purchase quantity and high cost, which increases the difficulty of the problem and make it more difficult to solve.

Raw material inventory control is one of the cores of production enterprises, especially one of the cores of enterprise logistics management for the continuous production [2-3]. Under the pressure of reducing the production cost continuously, the managers of iron and steel enterprise have to focus on inventory. Although inventory can make up for the damage caused by the uncertain factors and maintain the continuity of production, but also increase the inventory cost of the enterprise and lead to a large backlog of funds at the same time. Therefore, how to meet the iron and steel enterprise capacity constraints to optimize the balance between inventory and production requirements is a problem worthy to be discussed. Appropriate inventory is necessary to ensure the normal continuous production of enterprise, too little and too much can cause unnecessary economic losses. Inventory is too large, which will need extra storage places and increase the stock fee and transportation fee so as to slow capital turnover occupied. It may reduce the quality of raw materials or make the material deterioration because of the change of storage condition and the passage of time; If too little inventory, it may cause work being held up for lack of materials,

* Corresponding author. Tel: +86-156-97181-020; fax: +86-027-68862283;
E-mail: wustxie@126.com

and even force to stop production, thus losing customers and sales opportunities, reducing profit or losing their entitlement. The main purpose of material inventory control is that, under the condition of normal and continuous production, inventory takes up the least amount of money and less procurement costs in order to achieve the minimum total inventory cost. How to maintain in the best of our inventory and purchase quantity is the research questions this chapter [4-5].

2. Characteristics and Consider Factors of Iron and Steel Enterprise Raw Material Inventory

This chapter aims at raw material inventory problem in the process of the continuous production of iron and steel enterprise. For this type of enterprise, features are as follows:

a. Source of raw materials is extensive and raw material in inventory is multifarious so that the determination of safety inventory is difficult;

b. The demand of raw material is very big, and the order in advance is longer than the other;

c. There are a lot of subcontractors to provide the raw material for continuous production enterprise so that the number of received goods is uncertain. There are three forms of the contract signed with enterprises: their annual contract, quarterly contract and monthly contract. If the contract be enforced strictly, the arrival amount of goods each month is certain, but under the current situation in our country, due to transportation problems or imperfect delivery system, some subcontractors may provide the goods beyond the contract amount the goods and some others may only perform the part of contract, or even not. Hence it is brought certain problems to determine the arrival amount of goods;

d. The inventory cost is high; the raw material takes up more funds;

e. The shortage of the goods is not allowed, certain insurance reserve is stored;

f. Parts of the raw materials have a shelf life.

The inventory problem of raw material has features, such as many varieties, large quantity, high cost and long order time and so on, these features increase the difficulty of the problem to make it more difficult to solve. Considering the characteristics above, I put forward an inventory optimization model to determine the best inventory level and the best purchasing for all kinds of raw materials in a planning period in order to realize the minimization of inventory cost. What the inventory optimization should solve is: how to determine the single batch purchase quantity makes the total cost at least about the purchasing cost and storage cost [6].

The scale of the raw material inventory has a great influence on the production coordination and capital flow, therefore, the size of the raw material inventory should be generally considered from three aspects:

a. The production: from the perspective of production, enough inventory of raw material is in the hope. So, when some unplanned orders make the production increase temporarily or purchasing encounters the unexpected accident and long order time can't purchase raw materials in time, if there is enough raw material inventory, production can be continued so that the forced shutdown caused losses owning to the shortage of stock. So from taking the production into consideration, of course, the more reserve of raw material is better;

b. The capital flow: inventory of raw materials accounts for most of flow capital, the raw material is stored in the warehouse to produce collateral damage, namely the holding cost. The fee includes: interest, the storage of raw material loss, warehouse workers wages, warehouse handling fee, depreciation, repair, ventilation, lighting, rent and other fees of warehouse. These costs increase as the quantity and time increases. Therefore, in liquidity ways, the storage of raw materials is less as far as possible;

c. The procurement: considering from procurement, raw material need various fees from order into the library, such as communication fee of purchasing department, travel expenses for acceptance, handling fee, warehouse department acceptance fee. These costs are mostly related to the number of orders, which will increase with the increase of order number. So from the aspects of order, the less number of orders is better, it also requires more quantity each time.

From different point of view, therefore, the requirement of storage material is different. In these factors, there is a contradiction between the order and the storage. Less batches and large number of orders means low purchasing cost and the high cost of storage, while more batches and small number of orders will make the low storage cost and the high cost of purchase. Therefore, how to obtain the balance of contradictions through the reasonable optimization and achieve the lowest total inventory cost is the key to the inventory control.

3. The Establishment of the Optimize Inventory Model

3.1. THE ESTABLISHMENT OF THE MATERIAL INVENTORY MODEL

This chapter studies the raw material inventory problem of continuous production process of iron and steel enterprise. Inventory control problem is an optimization problem with multiple variables and complex constraints, which is difficult to solve through the traditional optimization method and is very suitable to solve by genetic algorithm. Genetic algorithm is a kind of adaptive random global search algorithm based on the biological evolution and natural genetic mechanism, which is suitable for solving complex problems such as global optimization, large-scale, multivariable, nonlinear. Inventory optimization unit is the core of the whole system; its main function obtains needed data from the database to optimize procurement and inventory with

certain optimization algorithm. The optimization results are applied to guide the procurement and inventory of raw materials and store a variety of data related to the inventory. Some of these data come from the actual production process, and some come from other optimization of logistics management and the output of forecasting subsystem [7].

In practical production, the comprehensive decision-making of inventory management is affected by various dynamic factors, such as macroeconomic factors, raw material and product market and the condition of the supplier [8]. If all these factors are taken into consideration, it will make the problem complex, and some of the values cannot be obtained directly. In order to simplify the problem, when optimizing raw material purchasing and inventory, we should mainly consider the production planning of manufacturer according to the sales situation and the actual production situation, the forecast value of raw material purchase prices in the coming year and inventory parameters prediction [9-10].

Inventory optimization model for the considerable scale enterprise, the total quantity of raw material purchasing and the total quantity of consumption monthly roughly equal [11-13]. Inventory model is set up in a month for a period with the monthly purchase quantity and inventory as control variable. Inventory optimization saves money to maximize, at the same time meet the needs of continuous production for the purpose.

The description of raw material inventory problem is as follows.

The objective function:

$$C_{\min} = \sum_{i=1}^{N} \left\{ k_i + C_i X_i + \frac{1}{2}\left[Y_i(j) + Y_i(j+1)\right] \cdot h_i \right\}. \quad (1)$$

Constraints:

a. The balance constraints of raw material

Balance is as follows before and after the period of raw material inventory:

$$X_i + Y_i(j) - Y_i(j+1) - \sum_{i \in R} (1+\delta) q_i = 0, \quad (2)$$

b. The balance constraints of product sale

Before and after the period, on the basis of considering inventory change, the same material meet the following relation on the quantity from raw material to finished product:

$$Y_i(j+1) = Y_i(j) + X_i - U_i, \quad (3)$$

c. Procurement constraints

The total procurement budget:

$$\sum_{i-1}^{N} C_i X_i \le C, \quad (4)$$

(In the type C is the total procurement budget.)

d. The matching constraint of ability to inventory

In actual production, it's hard to maintain completely the material inventory within the scope of the safety stock level, and it is often more than safety stock, so the matching of inventory capacity meets the following constraints:

$$S_i \le Y_i \le X_i, \quad (5)$$

e. The constraint of total supply capacity

The amount of material purchase should not be larger than the biggest supply of material in the period of time:

$$X_i \le Z_i, \quad (6)$$

k_i is the fixed startup cost of raw material i purchasing;

C_i is the procurement price of raw material i in month j;

X_i is the purchasing quantity of raw material i in month j;

$Y_i(j)$ is the inventory quantity of raw material i at the beginning of month j;

$Y_i(j+1)$ is the inventory quantity of raw material i at the end of month j;

U_i is the consumption of raw material i in month j;

h_i is the storage cost of unit goods of raw material i;

N is the kinds of raw materials;

S_i is the safety inventory of raw material i;

δ is the average coefficient of waste;

q_i is the number of raw material for production products.

3.2. THE SIMPLIFICATION OF MATERIAL INVENTORY MODEL

TABLE 1. The relevant parameters and values of calculation model

Parameter	Material		
	Iron ore	*Llimestone*	*Coal*
The unit price (RMB/t)	730	60	541
Unit inventory cost (RMB /ton)	14.0	5.0	18
Safety inventory (t)	50000	50000	50000
Monthly consumption (t)	444882	102851	5554
The maximum annual material supply (t)		3000000	2200000

$$C_{\min} = 737X_1 + 62.5X_2 + 550X_3 + 14Y_1 + 5Y_2 + 18Y_3 - 3415287.5 . \quad (7)$$

Constraints:

$$X_1 \geq 2Z = 340000$$
$$0 \leq X_2 \leq 250000$$
$$0 \leq X_3 \leq 183333$$
$$50000 \leq Y_2 \leq X_1$$
$$50000 \leq Y_2 \leq X_2$$
$$50000 \leq Y_3 \leq X_3$$
$$730X_1 + 60X_2 + 541X_3 \leq 100514224$$

3.3. THE SOLUTION OF MATERIAL INVENTORY MODEL

Recently, Genetic Algorithm (GA) is widely applied to the optimization problems in different fields, which arouses more and more people's interest of study and application due to its good characteristics that the Genetic Algorithm is not dependent on the problem model, along with global optimality, implicit parallelism, high efficiency and solution of the nonlinear problem.

GA is applied to the problem of constrained optimization; the constraint processing has become a very important link [14]. There are two kinds of solutions: one adopts the changed operator in the operation, producing always the legitimate offspring from the legal parent satisfying the constraint conditions and making the search in the legal and valid space all the time; Another is to apply the corresponding punishment to the objective function in the adaptive function according to each individual to satisfy the different constraints so that the constrained problems are converted into unconstrained problems. The main steps of genetic algorithm are as follows [15]:

a. Establish the initial group composed of a string at random;

b. Calculate the fitness of each individual.

c. According to the genetic probability, the following operations is applied to generate new group:

Copy: The existing excellent individual copied is added to the new group, and the bad individual is removed;

Hybridization: Two individuals picked out are exchanged, and the new individual is added to new groups;

Variation: One particular character is changed randomly in an individual; the new individual is added into new groups;

d. The steps of (2), (3) are carried out repeatedly to reach the termination conditions, and the best individual is chosen as a result of the genetic algorithm. It is the key to practical application to choose the encoding strategy and convert the parameters into a string and the accurate fitness. The initial population $N = 200$, the crossover probability $P_c = 0.95$, the mutation probability $P_m = 0.01$, the function of minimum inventory cost is as the objective function, namely $F(X) = C_{\min}$, the objective function is as fitness function directly, namely. The optimal purchase quantity and inventory is obtained in the following table with iterations of 53 times.

TABLE 2. The results of optimization model

	X_1	X_2	X_3	Y_1	Y_2	Y_3
The optimal value (t)	339930	50340	50550	50260	50270	50250

$$C_{\min} = 2.7947e + 008 , \text{(RMB)}. \quad (8)$$

The inventory of raw material is one of the cores of logistics management, on the one hand, the raw material inventory takes up a lot of money; on the other hand, the reasonable inventory is essential. In the process of enterprise management, Inventory items take up a lot of money. Iron and steel enterprises cost 20% to 40% of the profits every year to maintain its all inventory, enterprises set out to the material inventory to reduce the production cost and improve enterprise's capital turnover and return for the survive and development in the fierce market competition. Inventory management has become the important link in the production and management of iron and steel enterprise, and raw materials inventory accounts for most of the iron and steel enterprise, so now it is the important content of the inventory management that many iron and steel enterprises have to reduce raw material inventory levels and benefit from it.

4. Conclusions

With analysing the characteristics of the raw materials inventory and other factors, the optimization model of raw material inventory of iron and steel enterprise is established. The optimal purchase quantity and inventory is obtained by optimization analysis for the main raw materials such as iron ore, limestone and coal, the result makes the raw material inventory take up the least amount of money. The optimization control model of raw material inventory, being put forward, is mainly for the iron and steel enterprises with continuous production process, can also be extended to other type control of raw material inventory of the enterprise.

Acknowledgments

This research reported in the paper is supported by Key Laboratory of Metallurgical Equipment and Control of Ministry of Education in Wuhan University of Science and Technology (2013A03).This support is greatly acknowledged.

References

[1] Anfilets S., Shuts V. (2012) The use of natural optimization algorithms for the implementation of adaptive control at the crossroad. In *'Reliability and Statistics in Transportation and Communication'*, Riga, Latvia, October 17–20, 2012 Riga: Transport and Telecommunication Institute, 227–233

[2] Ji Feng, Zhang Jian-Min (2009) *Journal of Mechanical & Electrical Engineering*, **26**(2), 62-69

[3] Fang Shuiliang, Yao Yanfei, Zhao Shikui (2011) *Journal of Mechanical & Electrical Engineering*, **28**(3), 269-274

[4] Zhang Qin, Yan Ming-Zhong, CAI Lan (2006) *Journal of Mechanical & Electrical Engineering*, **23**(4), 62-69

[5] Wang Zhongzhou (2013) *Journal of Mechanical & Electrical Engineering*, **30**(1), 69-72

[6] Xue Changjiang (2007) The Steel Supply Chain of Forging Stability - Re-recognition of Current Relationship for Steel and Steel Production. *China's steel industry*, **1**, 33-35.

[7] Wang Wei, Wu Min, Chen Xiao-Fang, Gui Wei-Hua (2003) Raw Material Stock Optimization System Based on Multiple Parallel Genetic Algorithm. *Control Engineering of China*, **10**(1), 33-37

[8] Liu Guo-Li, Tang Li-Xin, Zhang Ming (2007) *Journal Of Northeastern University (Natural Science)*, **28**(2), 172-175

[9] Gao Peng, Liu Hao-Ran, Hao Xiao-Chen, Guo Feng, Shi Xin (2011) *Journal of Mechanical & Electrical Engineering*, **28**(2), 231-4

[10] Guo H, He J. (2013) *Computer Modelling and New Technologies*, **17**(3), 63-8

[11] Krainyukov A., Kutev V., Opolchenov D. (2010) *Transport and Telecommunication*, **11**(4), 14–28

[12] Peshkova I., Santalova D. (2011) On an optimal inventory problem for multi-echelon production system. *In the Proceedings of the 11th International Conference 'Reliability and Statistics in Transportation and Communication'*, Riga, Latvia, October 19-22, 2011. Riga: Transport and Telecommunication Institute, 249-254

[13] Guseynov Sh. E., Medvedev A. N., Baranova L. V. (2013) Mathematic modelling of the optimal cargo Handling management process at the large Transport hubs. *In the Proceedings of the 13th International Conference 'Reliability and Statistics in Transportation and Communication'*, Riga, Latvia, October 16–19, 2013. Riga: Transport and Telecommunication Institute, 188–196

[14] Zadachyn V., Dorokhov O. (2012) *Transport and Telecommunication*, **13**(4), 303–9

[15] Liang Ji-Ye (1999) *Computer application research*, **16**(7), 20-21

Authors

Yuanmin Xie, born November, 1972, China

Current position, grades: Master, Lecturer. College of Machinery and Automation, Wuhan University of Science and Technology

Scientific interest: Inventory management, Mechanical design and application

Publications: 10

An update algorithm of decision rules in expert systems based on rough sets theory

Yehong Han[1*], Lin Du[1]

[1] *Department of Computer, Qilu Normal University, Jinan, CHINA*

Received 1 January 2014, www.tsi.lv

Abstract

The decision table in rough sets theory is a kind of prescription, which specifies what actions should be undertaken when some of conditions are satisfied. Therefore, this tool can be used as knowledge representation system in expert systems. Decision rules, which are obtained by simplification of decision tables, can be used as rationale of decision reasoning. In order to compute new decision rules on the decision table in which a new instance is added, new instances are classified three cases according to the relation between the new instance and the original set of decision rules in the paper, and the category is proved that it is a partition of new instances. According to the category, an update algorithm of decision rules based on rough sets theory in expert systems is presented, and the complexity of the algorithm is obtained.

Keywords: rough sets, expert systems, incremental learning, decision table, algorithm

1. Introduction

Expert systems [1] are computer intelligence systems, which can execute special tasks as experts. In other words, expert systems are that expertise is transferred from a human to a computer, and computers can be used as a human consultant. In order to make the computer gives advices and performances like an expert, some techniques must be employed. Now expert systems can provide very powerful and flexible methods for obtaining solutions to a lot of different problems that often cannot be dealt with by other, more traditional methods [2, 3]. Although many scholars have studied various methodologies for knowledge processing in expert systems, and the technology of expert systems has made great progress, there are still many problems, for example, most of research on expert systems limit to static data, and neglect update algorithm of knowledge bases in expert systems [4].

Rough sets theory [5] is a mathematical tool to data analysis. It was presented by Zdzislaw Pawlak in 1982. It can be used to deal with fuzzy and uncertainty information. Now, rough sets theory have been widely used in a variety of domains, such as decision support system, machine learning, expert systems, pattern recognition and others [6]. According to rough sets theory, a decision table is a kind of prescription, which specifies what actions should be undertaken when some of conditions are satisfied. Therefore, this tool especially suitable for expert systems and it can be used as knowledge representation system [7] of expert systems. Decision rules [8], which are obtained by simplification of decision tables, can be used as rationale of decision reasoning.

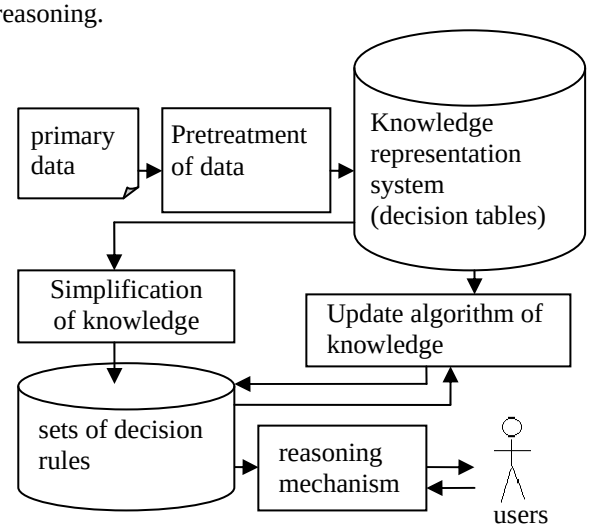


FIGURE 1. The model of expert systems based on rough

In the paper, a basic model of expert systems based on rough set theory is given as shown in figure1 by improving the traditional model. In order to improve the defect that the research of data in knowledge representation database of expert systems limit to static data, an update algorithm of knowledge based on the model of expert systems is presented. The new instance can be added to knowledge database of expert systems, and decision rules, which can be used as rationale of decision reasoning in expert systems will be updated by using the algorithm, and the algorithm is proved that it

* Corresponding author Tel: 13793161609
E-mail: sdzzhyh@163.com

can be used to consistent and inconsistent decision tables, and the complexity of the algorithm is given.

2. Rough Sets Theory

2.1. NOTATIONS AND DEFINITIONS OF ROUGH SETS

Knowledge representation system [9] is a pair $S =< U, A >$, where U is a nonempty and finite set called the universe, and A is a nonempty, finite set of primitive attributes. Every primitive attribute $a \in A$ is a total function $a: U \to Va$, where Va is the set of value of a, called the domain of a. Let $C, D \subset A$ be two subsets of attributes, called condition and decision attributes respectively. KR - system with distinguished condition and decision attributes will be called a decision table, and will be denoted as $S =< U, A >$.

Let P and Q are subsets of A. By P - Positive region of Q denoted $POS_F(Q)$, the set $POS_F(Q)$ can be computed by equation (1).

$$POS_F(Q) = \cup_{X \in U/IND(Q)} P\_(X), \tag{1}$$

where

$$IND(Q) = \{(x, y) \in U^2 \mid f(x,b) = f(y,b), x \in U, y \in U, b \in Q\}.$$

The family $R \subset C$ will be called a D - reduction of C, If and only if R is the D -independent subfamily of C and $POS_R(D) = POS_C(D)$.

In fact, the decision table can be viewed as a model for a set of propositions about reality, called here decision logic, which will be used to drive conclusions from data available in the Knowledge representation system. In decision logic language, there are definitions as follow:

The set of formulas in decision logic language is at least set satisfying the following conditions:

(1) Expression of the form (a, v) or in short a_v, called elementary formulas, are formulas of the DL - language for any $a \in A$ and $v \in V_a$.

(2) If θ, ψ are formulas of the DL - language, then so are $\neg\theta$, $\theta \vee \psi$, $\theta \wedge \psi$, $\theta \to \psi$ and $\theta \equiv \psi$.

An object $x \in U$ satisfies a formula θ in $S =< U, A >$, denoted $|x| =_s \theta$. Let $P = \{a_1, a_2, ..., a_n\}$, and $P \subseteq A$, formula of the form $(a_1, v_1) \wedge (a_2, v_2) \wedge ... \wedge (a_n, v_n)$ will be called a P - basic formula. If θ is a P -basic formula and $R \subseteq P$, then by θ / R we mean the R -basic formula obtained from the formula θ by removing from θ all elementary formulas (a, v_a) such that $a \in P - R$.

In decision logic language, any implication $\theta \to \psi$ will be called a decision rule. θ and ψ are referred to as the predecessor and the successor of $\theta \to \psi$ respectively. If a decision rule $\theta \to \psi$ is true in S, we will say that the decision rule is consistent in S. If $\theta \to \psi$ is a decision rule, where θ and ψ are P -basic and Q - basic formulas respectively, then the decision rule will be called a PQ - basic decision rule, (in short PQ - rule). Any finite set of decision rules will be called a decision algorithm. The decision algorithm is consistent in S, if and only if all its decision rules are consistent in S. If all decision rules in a decision algorithm are PQ - basic decision rules, the algorithm is said to be PQ - decision algorithm, or in short PQ - algorithm and will be denoted as (P, Q).

2.2. SIMPLIFICATION OF DECISION TABLES

In [10], Zdzislaw Pawlak proposed that, in order to simplify a decision table, three steps as follows should be taken:

1) Reduce the set of attributes, i.e. remove all superfluous columns from decision table.

2) Simplify the decision rules, i.e. eliminate the unnecessary conditions in each rule of the algorithm separately.

3) Remove all duplicate decision rules from the algorithm.

Example 1

Supposing $S =< U, A >$, where $U = \{1,2,3,4,5,6,7,8\}$ and $A = \{a, b, c, d, e\}$. $C = \{a, b, c, d\}$ and $D = \{e\}$ are condition and decision attributes respectively. The decision table is shown in table 1.

TABLE 1. The decision table of example 1

U	a	b	c	d	e
1	1	0	0	1	1
2	1	0	0	0	1
3	0	0	0	0	0
4	1	1	0	1	0
5	1	0	0	1	0
6	1	1	0	2	2
7	2	2	0	2	2
8	2	2	2	2	2

It is easy to compute that the only e-dispensable condition attribute is c. Hence there is a D-reduction of the family C, and $R = \{a, b, d\}$ [11, 12].

The next step is to simplify the decision rules. In [10], only consistent decision rules will be reduced. In fact, inconsistent decision rule can be reduced too. In order to simplify both consistent and inconsistent decision rules, we give the following definitions.

Let $\theta \to \psi$ is a decision rule in S, $\overline{|\theta \to \psi|_s}$ are referred to as the set of all objects which corresponding decision rules' predecessor are the same as θ, and its successor are not the same as ψ. Obviously, if the decision rule is consistent, $\overline{|\theta \to \psi|_s} = \phi$. Otherwise $\overline{|\theta \to \psi|_s} \neq \phi$.

Let $\theta \to \psi$ is a RD-rule, and $a \in R$. We will say that attribute 'a' is dispensable in the rule $\theta \to \psi$ if and only if $\overline{|\theta \to \psi|_s} = \overline{|\theta / R - \{a\} \to \psi|_s}$. Otherwise, attribute '$a$' is indispensable.

If all attribute $a \in R$ are indispensable in $\theta \to \psi$, then $\theta \to \psi$ will be called independent.

The subset of attributes $R' \subseteq R$ will be called a reduction of RD - rule $\theta \to \psi$, when $\theta / R' \to \psi$ is independent and $\overline{|\theta \to \psi|_s} = \overline{|\theta / R' \to \psi|_s}$.

If R' is a reduction of the RD-rule $\theta \to \psi$, then $\theta / R' \to \psi$ is said to be reduced.

According to the above definitions, the reductions of each decision rule in the algorithm [13] as shown in Table 2 will be obtained.

TABLE 2. The reduction of table 1

U	a	b	d	e
1	*	0	1	1
2	1	*	0	1
3	0	*	*	0
4	*	1	1	0
5	*	0	1	0
6	*	*	2	2
7	2	*	*	2
7	*	2	*	2
7	*	*	2	2
8	2	*	*	2
8	*	2	*	2
8	*	*	2	2

Remove all superfluous decision rules from the Table 2, the following minimum set of decision rules will be got [14].

$b_0 d_1 \to e_1$ from rule 1

$a_1 d_0 \to e_1$ from rule 2

$a_0 \to e_0$ from rule 3

$b_1 d_1 \to e_0$ from rule 4

$b_0 d_1 \to e_0$ from rule 5

$d_2 \to e_2$ from rules 6, 7 and 8

The set of decision rules, which specifies what actions should be undertaken when some of conditions are satisfied is particularly useful in decision making of expert systems. It can be used as rationale of decision reasoning [15, 16, 17, 18].

Most of research on expert systems based on rough sets limit to static data. In other words, the set of instances U in knowledge representation system is constant and unchanged, which will be called Closed World Assumption. But in many real life situations however this is not the case, and new instances can be added to the set U. This situation will be called the Open World Assumption [19]. In order to compute the minimum set of rules of decision table when a new instance is added, all the data in the decision table should be recalculated in the classical method [20, 21, 22]. Obviously, this method is not effective. In the paper, an update algorithm of decision rules will be introduced in the following section.

3. An Update Algorithm of Decision Rules Based on Rough Sets Theory

3.1. THE CATEGORY OF NEW INSTANCES

In order to introduce the algorithm, new instances, which will be added to decision tables in expert systems should be classified according to the relationship between the new instance and the old decision table.

Let $S = (U, A)$ is a decision table, M is a minimum set of decision rules before added the new instance. The new instance is *x*, and its *CD*-basic rule is $\theta_x \to \psi_x$. On the premise, there are the following definitions.

If there is a rule or several rules $\theta \to \psi$ in the set of decision rules M satisfying $\theta_x \to \theta$, and every rule satisfying $\theta_x \to \theta$ implies $\psi_x \equiv \psi$, then this situation will be called *x* matches M.

If there is a rule or several rules $\theta \to \psi$ satisfying $\theta_x \to \theta$ in M, and every rule satisfying $\theta_x \to \theta$ implies $\psi_x \neq \psi$, then this situation will be called *x* is totally in contradiction with M.

A new instance is partially in contradiction with M when there are several rules $\theta \to \psi$ satisfying $\theta_x \to \theta$ in M, and not only exist rules satisfying $\theta_x \to \theta$ satisfies $\psi_x \equiv \psi$, but also exist rules satisfying $\theta_x \to \theta$ satisfies $\psi_x \neq \psi$ in M.

Both the totally and partially contradiction will be called *x* is in contradiction with M.

If there is not any rule $\theta \to \psi$ satisfying $\theta_x \to \theta$ in a minimum set of decision rules, then we will say the new instance *x* is completely new in M.

According to above definitions, new instances will be classified three cases according to the relationship between the new instance and the old decision table.

1) *x* matches M;

2) *x* is in contradiction with M;

3) *x* is completely new in M.

This classification is completely, and covered all situations that new instances may be, in other words, it is a partition of new instances.

Suppose a new instance's *CD* -basic rule is $\theta_x \to \psi_x$ .The set of all objects which corresponding decision rules $\theta \to \psi$ satisfy $\theta_x \to \theta$ and $\psi_x \neq \psi$ will be called the contradicting domain of the new instance x, and will be denoted Cx , and Cx can be obtained by equation (2).

$$Cx = \left\{ i \middle| i \in U, \forall \theta_i \to \psi_i \in M \ satisfy \ \theta_x \to \theta_i \ and \ \psi_x \neq \psi_i \right\} \quad (2)$$

The set of all the objects which corresponding decision rules $\theta \to \psi$ satisfy $\theta_x \to \theta$ and $\psi_x \equiv \psi$ will be called the matching domain of the new instance x, and will be denoted Mx , and Mx can be obtained by equation (3).

$$Mx = \left\{ i \middle| i \in U, \forall \theta_i \to \psi_i \in M \ satisfy \ \theta_x \to \theta_i \ and \ \psi_x \neq \psi_i \right\} \quad (3)$$

The rule $\theta_i \to \psi_i$ is referred to as the reduction of the corresponding rule of object i . Let $S = (U, A)$ is a decision table, and there are not same rows in the decision table. R is a reduction of condition attributes. M is a minimum set of decision rules before adding a new instance. A new instance is *x*, and let its *CD* - basic rule be $\theta_x \to \psi_x$. The new decision table will be denoted $\theta_x \to \psi_x$ after adding the new instance, where $U' = U \bigcup \{x\}$. By the above definitions, the following proposition hold.

Proposition 1

If for every object $y \in Cx$ does not imply $y \models_S \theta_x / R$ in S , a reduction of condition attributes of the new decision table $S' = (U', A)$ is R .

Proof: Because there is no an object $y \in Cx$ such that $y \models_S \theta_x / R$ in S , the new instance is not in contradiction with all the *RD* - basic rules including in S . When the new instance x is added to S , $POS_R(D)' = POS_R(D) \bigcup \{x\}$ and $POS_C(D)' = POS_C(D) \bigcup \{x\}$, where $POS_R(D)'$ and $POS_C(D)'$ are R - Positive region and C - Positive region of D respectively. Because $POS_C(D) = POS_R(D)$, $POS_C(D)' = POS_R(D)'$ is true, and R is the D - independent subfamily of C in universe U' , a reduction of condition attributes of the new decision table $S' = (U', A)$ is R .

Proposition 2

If there is an object $y \in Mx$ implies $y \models_S \theta_x$ in S , then the decision table $S' = S$, and its reduction of attributes and minimum set of rules will not change.

Proof: there is a object $y \in Mx$ satisfying $y \models_S \theta_x$ in S , then the new instance have existed in the decision table. Hence, the decision table does not change when we add the new instance to it. Obviously, the reduction of condition attributes and the minimum set of rules of decision table S' will not change.

Proposition 3

If there is only one object $y \in Cx$ such that $y \models_S \theta_x$, and there is not any other object $i \in Cx$ and $i \neq y$ satisfying θ_x / R , then there is a reduction R' of condition attributes satisfying $R' \subseteq R$ of $S' = (U', A)$.

Proof: If there is only one object $y \in Cx$ satisfying $y \models_S \theta_x$ in S, then the new instance x will be in contradiction with the object satisfying θ_x . There is not any other object $i \in Cx$ satisfying θ_x / R , then $POS_R(D)' = POS_R(D) - \{ y \mid y \models_S \theta_x \}$ and $POS_C(D)' = POS_C(D) - \{ y \mid y \models_S \theta_x \}$. Hence, we will conclude that $POS_C(D)' = POS_R(D)'$. Because the number of elements of the positive region is reduced when the new instance is added to S, R may be not the D-independent subfamily of C in universe U'. There is a reduction R' of condition attributes satisfying $R' \subseteq R$ of decision table $S' = (U', A)$.

Proposition 4

If there are two objects $y \in Cx$ satisfying $y \models_S \theta_x$ at least, then a reduction of attributes of decision table $S' = (U', A)$ is R .

Proof: there are two objects $y \in Cx$ satisfying $y \models_S \theta_x$ at least, then $POS_R(D)' = POS_R(D)$ and $POS_C(D)' = POS_C(D)$. Hence, $POS_C(D)' = POS_R(D)'$. Because the positive region doesn't change after adding the new instance, R is the D-independent subfamily of C in universe U'. So a reduction of attributes of decision table $S' = (U', A)$ is R .

Proposition 5

Let $\theta_i \to \psi_i$ is a reduction of the corresponding *CD* - rule of the object i in S , and $i \notin Cx \bigcup \{ j \mid \theta_j' \equiv \theta_i' , i \in Cx,$ and $j \in Mx,$ where θ_i' and θ_j' are corresponding C – basic formulas of objects i and j

respectively}. Let W be the set of all minimum set of rules that can be obtained by reducing S'. If a reduction of attributes doesn't change when a new instance is added to S, there is a minimum set of rules $M' \in W$ satisfying $\theta_i \to \psi_i \in M'$.

Proof: because $i \notin Cx \cup \{ j \mid \theta_j' \equiv \theta_i', \ i \in Cx, \text{ and } j \in Mx, \text{ where } \theta_i' \text{ and } \theta_j' \text{ are corresponding } C \text{ - basic formulas of objects } i \text{ and } j \text{ respectively}\}$, obviously, the rule $\theta_x \to \psi_x$ is not in contradiction with the decision algorithm composed by rules satisfying $\theta_i \to \psi_i$. Hence there is a minimum set of rules $M' \in W$ satisfying $\theta_i \to \psi_i \in M'$.

According to above propositions, following corollaries will be established.

1) If the new instance x matches M, a minimum set of decision rules of decision table S' is still M.

2) If the new instance x is completely new in M, a reduction of condition attributes of S' is still R, and a minimum set of rules, which can be denoted M' can be computed by equation (4),

$$M' = M \cup \left\{ \theta_x' \to \psi_x \middle| \theta_x' \to \psi_x \text{ is a reduction of } \theta_x \to \psi_x \right\} \quad (4)$$

3) if a reduction of attributes doesn't change after adding a new instance, in order to compute the minimum set of rules, we need simplify corresponding RD -rules of objects in $Cx \cup \{ j \mid \theta_j' \equiv \theta_i', \ i \in Cx, \text{ and } j \in Mx, \text{ where } \theta_i' \text{ and } \theta_j' \text{ are corresponding } C \text{ -basic formulas of objects } i \text{ and } j \text{ respectively}\} \cup \{x\}$.

According to above propositions and corollaries, the following method shown in figure 2 can be used to compute the minimum set of rules. And the method can be used in the update algorithm of decision rules in expert systems.

3.2. AN UPDATE ALGORITHM OF DECISION RULES

Input: $S = (U, A)$ (S is a decision table);

R (R is a reduction of attributes of S);

M (M is a minimum set of rules of decision table S based on the attribute reduction R);

$\theta_x \to \psi_x$ ($\theta_x \to \psi_x$ is the new instance's CD-basic Rule).

Output: M' (M' is a Minimum Set of Rules of Decision Table $S' = (U' = U \cup \{x\}, A)$).

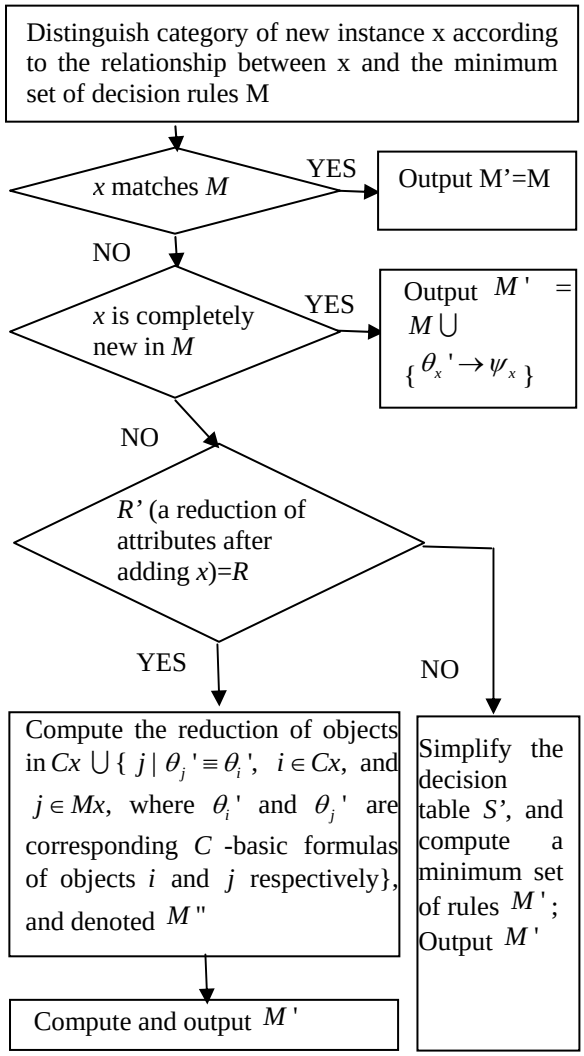


FIGURE 2. Algorithm to simplify decision tables

Algorithm:

Step1: $M1 = M2 = M3 = \phi$;

Step2: FOR every rule $\theta \to \psi$ in M

IF $\theta_x \to \theta$ is not true

THEN $M1 = M1 \cup \{\theta \to \psi\}$

ELSE IF $\psi_x \equiv \psi$

THEN $M2 = M2 \cup \{\theta \to \psi\}$

ELSE $M3 = M3 \cup \{\theta \to \psi\}$

Step3.1:

IF $M2 = M3 = \phi$

THEN {

Compute reductions of the rule $\theta_x \to \psi_x$,

and a result is denoted as $\theta_x' \to \psi_x$;

$M' = M \cup \{\theta_x' \to \psi_x\}$;

}

Step3.2:

ELSE IF $M3 = \phi$

THEN $M' = M$;

Step3.3:

ELSE {

Step3.3.1: Compute the contradicting domain of x:

$$Cx = \left\{ \begin{matrix} i \mid i \ \ is \ \ an \ \ object \ \ which \\ corresponding \ \ rule \ \ is \ \ in M3 \end{matrix} \right\};$$

Step3.3.2: Compute the matching domain of x:

$$Mx = \left\{ \begin{matrix} i \mid i \ \ is \ \ an \ \ object \ \ which \\ corresponding \ \ rule \ \ is \ \ in M2 \end{matrix} \right\};$$

Step3.3.3: IF $\exists y \in Mx$ satisfying $y \models_S \theta_x$

THEN $M' = M$

Step3.4:

ELSE {

Step3.4.1 Compute the set:

$C_{Cx} = \{\theta_j \to \psi_j \mid \theta_j \equiv \theta_i, i \in Cx, j \in Mx\}$;

(Where $\theta_j \to \psi_j$ is the corresponding rule of object j)

IF do not exist $y \in Cx$

Satisfying $y \models_S \theta_x / R$

THEN $R' = \{R\}$

ELSE IF at least exist two objects

$y \in Cx$ Satisfying $y \models_S \theta_x$

THEN $R' = \{R\}$

ELSE IF only exist an object

$y \in Cx$ satisfying $y \models_S \theta_x$

THEN compute the set of all The D-reductions of R in U' and denoted R'

ELSE compute the set of all the D-reductions of C in U' and denoted R';

Step3.4.2 IF $R \in R'$

THEN {

Compute the reduction of Corresponding decision rules of objects in $Cx \cup \{ j \mid \theta_j' \equiv \theta_i', i \in Cx,$ and $j \in Mx,$ where θ_i' and θ_j' are corresponding C - basic formulas of objects i and j respectively$\} \cup \{x\}$, and denoted M'';

$M' = M'' \cup M1 \cup M2 - C_{C_x}$;

}

Step3.5

ELSE {

Simply all the decision rules in S' according to a reduction of attributes which have computed, and obtain a minimum set of rules M'

}

}

}

Step 4: OUTPUT M'.

Let m and n be referred to as the quantity of elements in the set of condition attributes and universe respectively. Obviously, if x matches M, or x is completely new in M, the time-complexity of the algorithm is $O(mn)$. If x is in contradiction with M and there is not an object $y \in Cx$ satisfying $y \models_S \theta_x / R$ or there are two objects $y \in Cx$ satisfying $y \models_S \theta_x$ at the same, the time-complexity of the algorithm is $O(mn^2)$. At those situations, the algorithm is better than the classical algorithm obviously. The time-complexity of the classical algorithm is $O(2^m n^2)$ [23, 24]. But in other situations, the efficiency of this algorithm is worse than the classical algorithm little, and its time-complexity is $O(2^m n^2)$ too.

3.3. EXAMPLES

In order to understand the algorithm, examples of decision rules update will be given as follows:

Example 2:

If a new instance $a_0 b_1 c_2 d_1 \to e_0$ will be added to the decision table as shown in table1. Because $a_0 b_1 c_2 d_1 \to a_0$, and $e_0 \equiv e_0$, the new instance matches the minimum set of decision rules obtained by example 1, According to above algorithm, the new minimum set of decision rules will be obtained as follows:

$b_0 d_1 \to e_1$	from rule 1
$a_1 d_0 \to e_1$	from rule 2
$a_0 \to e_0$	from rule 3, 9
$b_1 d_1 \to e_0$	from rule 4
$b_0 d_1 \to e_0$	from rule 5
$b_0 d_1 \to e_0$	from rule 6, 7, 8

Example 3:

If the new instance is $a_0 b_0 c_1 d_1 \to e_1$, the new minimum set of decision rules is as follows:

* $a_1 b_0 d_1 \to e_1$	from rule 1
$a_1 d_0 \to e_1$	from rule 2
* $a_0 d_0 \to e_0$	from rule 3
$b_1 d_1 \to e_0$	from rule 4
* $a_1 b_0 d_1 \to e_0$	from rule 5
$d_2 \to e_2$	from rule 6, 7, 8
* $a_0 d_1 \to e_1$	from rule 9

Because the new instance is in contradiction with M, and there is not an object $y \in Cx$ satisfying $y \models_S \theta_x / R$, the reduction of corresponding decision rules of objects in $Cx \cup \{ j \mid \theta_j' \equiv \theta_i', i \in Cx,$ and $j \in Mx,$ where θ_i'

and θ_j' are corresponding C-basic formulas of objects i and j respectively$\}\bigcup\{x\}$ should to be computed only.

Example 4:

If the new instance is $a_1b_2c_1d_1 \to e_2$, the new instance is completely new in M. Because If the new instance is completely new in M, a reduction of condition attributes of S' is still R, and a minimum set of rules is M', and $M' = M \bigcup \left\{\theta_x' \to \psi_x \middle| \theta_x' \to \psi_x \, is \, a \, reduction \, of \, \theta_x \to \psi_x\right\}$, the minimum set of decision rules is as follows:.

$b_0d_1 \to e_1$ from rule 1
$a_1d_0 \to e_1$ from rule 2
$a_0 \to e_0$ from rule 3
$b_1d_1 \to e_0$ from rule 4
$b_0d_1 \to e_0$ from rule 5
$d_2 \to e_2$ from rule 6, 7, 8
* $b_2 \to e_2$ from rule 9

4. Conclusions

In the paper, decision tables are used as knowledge representation in expert systems and the minimum set of rules are used as rationale of decision reasoning, based on this, a basic model of expert systems based on rough set theory is given. And the update algorithm of knowledge representation system in expert systems is discussed in detail. In order to compute new decision rules, new instances, which will be added to decision tables in expert systems are classified three cases according to the relation between the new instance and the minimal set of decision rule that have been computed. The category is a partition of all new instances. According to the category, an update algorithm of decision rules is presented. The algorithm can be used in a variety of other domains, such as machine learning, data analysis and so on. But when the new instance is in contradiction with M and the reduction of condition attributes is changed after adding the new instance, all the data must be recalculated. This situation must be researched forward.

Acknowledgements

This work was supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2013FL010).

References

[1] Russell S. J., Norvig P. (1995) *Artificial intelligence: a modem approach*, Prentice hall
[2] Jackson P. (1998) *Introduction to Expert Systems* Boston, MA, USA: Addison-wesley Longman publishing Co Inc.
[3] Jonassen, D. H., Wilson B. G., Wang S. and Grabinger R. S. (1993) *Journal of Computer Based Instruction*, **20**(3), 86-94
[4] Siler W., Buckley J. J. (2005) *Fuzzy Expert systems and Fuzzy Reasoning*. John Wiley & Sons Inc.
[5] Pawlak Z. (1991) *ROUGH SETS: Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers
[6] Wang Guo-Yin, Yao Yi-Yu, Yu Hong (2009) A Survey on Rough Set Theory and Applications. *Chinese Journal of Computers*, **7**, 1229-46
[7] Zdzislaw Pawlak (2002) *Information Sciences*, **147**(1-4), 1-12
[8] Yunliang Jiang, Congfu Xu, Jin Gou, Zuxin Li (2004) Research on Rough Set Theory Extension and Rough Reasoning. *2004 IEEE International Conference on Systems, Man and Cybernetics*, 5888-93
[9] Hong Shi, Jin-Zong Fu (2005) A Global Discretization Method Based on Rough Sets. *Machine Learning and Cybernetics*, 3053-57
[10] Zdzislaw Pawlak, Andrzej Skowron (2007) *Information Sciences*, **177**, 3-27
[11] HU Qing-Hua, YU Da-Ren, XIE Zong-Xia (2008) *Journal of Software*, **19**(3), 640-649
[12] Xiaodong Liu, Witold Pedrycz (2009) *IEEE Transactions On Knowledge And Data Engineering*, **21**(3), 443-462
[13] Thangavel K., Pethalakshmi A. (2009) *Applied soft computing*, **9**(1), 1-12
[14] Yee Leung, Wei-Zhi Wu, Wen-Xiu Zhang (2006) *European Journal of Operational Research*, **168,** 164-80
[15] Huang B., Li H., Wei D. (2012) *Knowledge-Based Systems*, **28,** 115-123
[16] Fazel Zarandi M. H., Rezaee B., Turksen I. B., Neshat E. (2009) *Expert Systems with Applications*, **36**(1), 139-154
[17] Yuhua Qian, Jiye Liang, Chuangyin Dang (2010) *IEEE*, **40**(2), 420-31
[18] Zdzislaw Pawlak (1997) *European Journal of Operational Research*, **99**(1), 48-57
[19] Jian Yu, Shuigeng Zhou, Lipo Wang, Jingsheng Lei (2009) *Computers & Mathematics with Applications*, **57**(6), 865-6
[20] Yunfei Yin, Guanghong Gong, Liang Han (2009) *Computers & Mathematics with Applications*, **57**(1), 117-26
[21] Qinghua Hu, Daren Yu, Maozu Guo (2010) *Information Sciences*, **180**(10), 2003-22
[22] Hong Yu, Dachun Yang, Hong Tang, Zhongfu Wu (2003) *Computer Engineering and Application*, **33,** 38-41
[23] Dongya Li, Baoqing Hu (2007) A Kind of Dynamic Rough Sets. *Fourth International Conference on Fuzzy Systems and Knowledge Discovery* (FSKD 2007)
[24] Herbert J. P., Jing Tao Yao (2009) *Computers & Mathematics with Applications*, **57**(6), 908-18

Authors

Yehong Han, born in 1981

Current position, grades: a lecturer with Department of Computer at Qilu Normal University
Scientific interest: M.S. degree in Management Science and Engineering from Shandong Normal University.
Scientific interest: data mining, data analysis, decision support systems and patter recognition

Lin Du, born in 1978

Current position, grades: an associate professor with Department of Computer at Qilu Normal University
University studies: M.S. degree in Computer Science and Technology from Shandong University
Scientific interest: software reengineering, system comprehension, software measurement and image retrieval

Authors' index

Cumulative Index

Nanoscience and Nanotechnology

N. Mykytenko, D. Fink, A. Kiv. Generalized Model of Pulsating Track Device
*Computer Modelling & New Technologies, 2014, **18**(1), 7-12*

A model for description and study of pulsating track-based devices is developed. The track electronics opens up perspectives for solving new scientific and technical problems. The successful solution of these problems requires an elucidation of the mechanisms of the functioning of track-based devices. In this paper, the nature of the pulsating behaviour of electric parameters of track devices is clarified using a specially developed model based on classical Molecular Dynamics. It is demonstrated that the model describes adequately the main features of pulsations in track devices that were established experimentally.

Keywords: track electronics, current pulsations, molecular dynamics

T. Lobanova-Shunina, Yu. Shunin. Nanothinking and Nanoeducation: Nanoscientific Literacy for Responsible Consumer Decision Making
*Computer Modelling & New Technologies, 2014, **18**(1), 13-24*

Consciously or unconsciously, the term 'nanotechnologies' is firmly entering the life of every consumer-citizen of the global community designating both relatively simple nanomaterials and goods that have already entered the market, and very complex technologies that are supposed to change radically the future of mankind. Applications based on today's basic research are expected to form the next industrial revolution. The unique properties of nanotechnology applications suggest potential to solve some of the most pressing social and business challenges, but they come with uncertainties and risks as all new technologies.
Taking advantage of technological progress and preventing adverse side-effects requires analysis, evaluation and guidance to ensure technology is developed in ways that benefits wider consumer society and every individual consumer on the planet. Unfortunately, general public lacks understanding and awareness of the basic properties, and sometimes even the existence of nanotechnologies and their implications linked to the consumption of nanoproducts. Moreover, a generally sceptical attitude among society groups prevails towards new technologies.
The general lack of public knowledge about nanoproducts that are already on the market in a full swing is likely to bring irrational and erroneous, potentially harmful, results. Therefore, modern technology requires educated work-force and responsible consumers and hence imperative for educated population.
Our mission had a focus on introducing changes into the curriculum to eliminate gaps in scientific knowledge of students (as potential consumers, managers and scientists) and to foster an active approach to developing responsible scientific consumption practices and to offer an opportunity for students from a wide range of disciplines to learn about nanoscience and nanotechnology, to explore these questions, and to reflect on the place of new technologies in the spheres of their major and in the global society.

Keywords: Nanotechnologies, responsible scientific consumption, consumer identities, nanoeducation, nanothinking

Mathematical and Computer Modelling

J. Zhou, W. Q .Ma, J. Y. Miao. Supplies Transportation Planning in Power Grid Urgent Repair Based on Hierarchical Genetic Algorithm
*Computer Modelling & New Technologies, 2014, **18**(1), 25-30*

To optimize supplies transportation solutions of power grid, a hierarchical genetic algorithm is put forward. Double hierarchical objective function is set by weighting, that is, when the precondition of inequality constraint is met, objective solution is to find out the shortest supply time and if the solution is also within the time constraint, the final solution would be the least supply cost, otherwise, it would still be the shortest supply time. The solution for optimal supplies transportation scheme at the least cost would be worked out by iteration of genetic algorithm. Compared with single objective genetic algorithm in simulation, hierarchical genetic algorithm is proved more effective and superior to decrease economic loss of accidents.

Keywords: Hierarchical Genetic Algorithm, Time constraint, Transportation Cost, Transportation Time, Economic Loss

Zengqiang Ma, Yacong Zheng, Sha Zhong, Xingxing Zou. Study of Maneuvering Target Tracking Algorithm Based on Kalman Filter and ANFIS
*Computer Modelling & New Technologies, 2014, **18**(1), 31-37*

Although Kalman filtering algorithm has been widely used in the maneuvering target tracking, conventional Kalman filtering algorithm always fails to track the maneuvering target as the target changes its movement state suddenly. In order to overcome its disadvantages, an improved Kalman filtering algorithm that based on the adaptive neural fuzzy inference system (ANFIS) is proposed in this paper. In the improved algorithm, the covariance matrix of Kalman residual is gainer and the measurement noise covariance can be updated in real-time by ANFIS module. Finally, the comparison and analysis of the experiment results between the original Kalman filtering algorithm and the improved one has been carried out. The experiment results show that the tracking error is obviously reduced and the accuracy is significantly boosted after the original Kalman filtering algorithm was substituted by the improved one.

Keywords: maneuvering Target tracking, Kalman filter, Adaptive Neuro-Fuzzy Inference System (ANFIS)

Ji Mingyu, Li Yanmei, Chen Zhiyuan. The Property Verification Methods of Complex Stochastic System Based on Directed Graph
*Computer Modelling & New Technologies,2014, **18**(1), 38-43*

With the increasing complexity of computer hardware and software systems, how to ensure system accuracy and reliability becomes an increasingly pressing issue. The quantitative verification of multiple until formula property has important practical significance in the field of biology. In this paper, for particular probability reward model, we give the detailed analysis of properties verification methods of the multiple until formula with transition step and transition reward constraints based on the weighted directed graph. At last, the example analysis is given. The theoretical analysis and example result show that the feasibility and validity of the method.

Keywords: Probabilistic system; Model checking; Multiple until formula; Directed graph

Zhang Yongheng, Zhang Feng. Research on the Smart Wireless Sensor Perception System and its Application Based on Internet of Things
*Computer Modelling & New Technologies, 2014, **18**(1), 44-51*

In order to solve the uncertain perception information appears in the perception process of intelligent wireless sensor, this paper considers the intelligent perception problem of Internet of Things (IoT) based on context perception. The current status of the research on intelligent perception and its existing problem is analyzed, and then a context perception method to solve the intelligent perception problem of Internet of Thing is proposed. The intelligent perception context description model of Internet of Things is constructed. In addition, it was investigated that how the intelligent routing maintained under fault conditions, and intelligent information management system of agriculture's was proposed of agricultural IoT system, combined with agricultural automatic control devices, which had already been successfully used in the agricultural production.

Keywords: Internet of Things, intelligent perception, automatic control, context perception, intelligent wireless sensor

Huang Li. Data Cleansing Base on Subgraph Comparison
*Computer Modelling & New Technologies 2014, **18**(1), 52-60*

With the quick development of the semantic web technology, RDF data explosion has become a challenging problem. Since RDF data are always from different resources, which may have overlap with each other, they could have duplicates. These duplicates may cause ambiguity and even error in reasoning. However, attentions are seldom paid to this problem. In this paper, we study the problem and give a solution, named K-radius sub graph comparison (KSC). The proposed method is based on RDF-Hierarchical Graph Model. KSC combines similar and comparison of 'context' to detect duplicate in RDF data. Experiments on publication datasets show that the proposed method is efficient in duplicate detection of RDF data. And KSC is simpler and less time-costs than other methods of graph comparison.

Keywords: RDF data cleansing, K-radius sub graph comparison

Li Gongfa, Liu Jia, Jiang Guozhang, Kong Jianyi, Xie Liangxi, Xiao Wentao, Zhang Yikun, Cheng Fuwei. Numerical Simulation of Flow, Temperature and Phase Fields in U71Mn Rail-Head Quenching Process
*Computer Modelling & New Technologies, 2014, **18**(1), 61-70*

With the sustainable and fast development of Chinese economy, the volume of railway freight was increasing, and which promoted the train load and speed continue improving, in order to meet the needs of high speed and over

loading trains' running, heavy and seamless steel rail has increasingly widely used Heat-treatment was emphasized for its important role to qualify the products of heavy rail. And air-cooling quenching was one of the widely used heating process methods. Air-jet is a very vital instrument in quenching by air-cooling. At present, the international community has been widely used air-cooled quenching, most quenching production lines of domestic converted into wind cold quenching line. It is necessary to simulate the inner and outer flow-field of air-jet. In this paper, by means of computational fluid dynamics soft system Fluent to establish the geometric model of heavy rail, analysis the distribution of the internal and external flow field about air-jet centre cross section and the three sections were in parallel with it. Then through setting surface heat transfer coefficient of air-cooling, numerical simulation of temperature field in the cooling process of heavy rail. Finally, the phase changing temperature of steel U71Mn was got based on its CCT curves. With the cooling curves of several key points, the cooling rate at phase transition point was calculated. By comparing with every microstructure's critical cooling rate, the final cooling microstructure was predicted. Relative tests showed that the prediction was reasonable. It is significantly valuable for parameters' selection in heavy rail's technical operation.

Keywords: temperature field, flow field, phase field, air-cooling, heavy rail

Information Technologies

Ranran Man, Dongsheng Zhou, Qiang Zhang . An Improved Collision Detection Algorithm Based on OBB
Computer Modelling & New Technologies, 2014, ***18****(1), 71-79*

In this paper, we present an effective algorithm based on the Oriented Bounding Box (OBB). Particularly, the article focus on how to reduce the amount of the time that the intersection of bounding boxes are detected for increasing the efficiency of collision detection algorithm through the following three steps. Firstly, the detection time of bounding box intersection is decreased by a pre-treatment method. Secondly, we optimize the traversal approach of the bounding box tree and reduce the depth of the two fork tree under take into account the temporal and spatial correlation of motion of virtual environment. Then, we reduce the breadth of tree result from this algorithm compares the distance between the bounding boxes and traverse the node of distance that near the tree. At last, the validity of the algorithm is verified by programming simulations.

Keywords: Collision detection; Oriented bounding box (OBB); Spatio-temporal correlation; Bounding volume hierarchy

Liu Hongxia, Zhang Feng. Precision Agriculture Compressed Sensing and Data Fusion Algorithm for Wireless Sensor Networks
Computer Modelling & New Technologies, 2014, ***18****(1), 80-84*

In order to improve the energy efficiency of WSN nodes and prolong the life of the network, reduce data redundancy. In this paper, proposed the spatial correlation node data compression and fusion algorithm based on the theory of compressed sensing. Firstly, make signal node random projection based on time correlation, then, for random routing instability and network transmission of data fusion technology reconstruction energy effects, proposed the energy consumption of compressed sensing and clustering data fusion technology. The experiment showed that after data fusion, not only effectively removes the redundant information of neighbouring nodes, and the reconstruction error is small, and can accurately realize data decompression, thereby reducing the node communication in wireless sensor network capacity, reduce power consumption of node, provides important support for the field environment in large scale wireless sensor network deployment.

Keywords: compressed sensing, intelligent wireless sensor, agriculture IoT, automatic control, data fusion

Yun-Feng Wang, Wei Cheng. Structural Design and Active Control Simulation for Double Beam Actuator Based on ANSYS
Computer Modelling & New Technologies, 2014, ***18****(1), 85-92*

A double-beam actuator (DBA) is designed in this study for attenuating the harmful vibration. Its finite element model including the folded beam, glue layers and piezoelectric laminar is developed in commercial software ANSYS. DBA's dynamic characters are analysed by employing modal and transient analysis in ANSYS. The calculated results are used to develop an explicit state space model by using the observer/Kalman filter identification (OKID) technique and Eigen-system Realization Algorithm (ERA). The robust H_2 controller is designed based on the identified state space model and it is then incorporated into the ANSYS finite element model to perform the close loop controlling simulation. The results of simulation show the settling time reduce to 0.06s with active control under impulse exciting

and the response amplitude decrease 20 dB under sine exciting, which demonstrate the validity of DBA in application of active control.

Keywords: active control, system identification, active control, ANSYS, piezoelectric

Weihua Yuan, Hong Wang. Using Two-Stage Non-Negative Matrix Factorization for Topic Recommendation in Online Social Networks

*Computer Modelling & New Technologies, 2014, **18**(1), 93-99*

This paper put forward a two-stage non-negative matrix factorization (TSNMF) for topic recommendation in online social network to solve the existing problems of mass data, extreme sparseness and cold start. In stage I, we use co-clustering to divide user-topic interest matrix into smaller sub-matrices called cluster-sub-matrices based on non-negative matrix factorization on interest-density matrix D. Each cluster-sub-matrix is much smaller than the original with similar internal interest pattern. In stage II, we use weighted non-negative matrix factorization algorithm to predict unknown items on each of cluster-sub-matrix directly. Experiments on real datasets show that TSNMF can not only gain high prediction accuracy on extreme sparse datasets, but avoid the problem of too much computation of NMF on the whole user-topic interest matrix, as well as the problem of the recommendation's quick local convergence.

Keywords: Recommender algorithm; NMF (non-negative matrix factorization); Clustering; Data sparseness

Operation research and decision making

D. Zagulova, R. Muhamedyev, I. Ualiyeva, A. Mansharipova, E. Muhamedyeva. Optimization of Medical Information Systems by Using Additional Factors

*Computer Modelling & New Technologies, 2014, **18**(1), 100-108*

Increasing longevity is one of the most important problems of modern Gerontology. Solution of these problems is connected principally with the use of information and communication technologies. Creation of a comprehensive health information system requires consideration of many factors, such as qualitative screening system based on patients' self-assessment, identification of possible errors that affect decision-making and patients' personal characteristics. The work presents the results of elderly Almaty and Almaty Region population survey conducted with the help of Active Longevity Portal designed for data collection, analysis and assistance to the elderly population of Kazakhstan. The results showed that the number of medical consultations is directly related to health self-assessment and anxiety levels. Detection of cardiovascular diseases (CVD) with the help of effort angina self-assessment demonstrated low sensitivity. Correlation between the Kettle's index of effort angina self-assessment, the impact of Physical Component Score (PCS) of SF-12 test onto the manifestation of cardiovascular disease in hereditary background, anxiety level and coronary heart disease manifestation, impact of Health Survey estimated by Physical Component Score (PCS) and Mental Component Score (MCS) SF- 12 test onto the correspondence between Effort Angina Questionnaire and CVD patient state was detected. Studies showed that detection of diseases through Questionnaire Survey self-assessment in certain situations may lead to significant errors. Consideration of these factors will help to build a more powerful information system in which personal data will be combined with clinical data and expert estimates.

Keywords: gerontology, Kazakhstan population, information technology, cardiovascular diseases, medical information system

M. Fay, V. Grekul, N. Korovkina. IT Investments Justification Based on the Business Driver Tree

*Computer Modelling & New Technologies, 2014, **18**(1), 109-114*

Choosing the right IT project for supporting the company business development is nowadays one of the most critical tasks in information technology management. No one has yet managed to create one optimal solution, equally suitable for different types of stakeholders (business owners, managers, investors). This article, based on the ideas of Value Based Management and business/value-driver trees, concentrates on an original approach to managing investments in enterprise architecture IT component. The method suggested has been successfully applied to evaluate the IT project portfolio within a large metals company in Russia.

Keywords: Value Based Management, IT-project, IT investments, business driver, IT projects portfolio, investments justification

Bang-Jun Wang. CBM-based Integrated Management Information System Design for Mine Construction Enterprises

*Computer Modelling & New Technologies, 2014, **18**(1), 115-120*

To solve existing problems in the management for mine construction enterprises, this research introduced the concept of CBM (Comprehensive Budget Management) on the basis of data and information demand analysis using enterprise management decision. Next, under the circumstance of project management, the market mechanism was introduced in mine construction enterprises, materializing the CBM-based integrated management information system design for mine construction enterprises. The system architecture encompassed six modules, namely, production progress management, project material management, mechanical and electrical equipment management, human resource management, integrated cost management and performance management. The system covered the integrated management information system for each process of the mine construction management. With the B/S structure, technological development approaches of this system consisted of UML modelling technique, dynamic configured technology, database design and implementation. The actual application of integrated management system in sample mine enterprises showed optimized enterprise management process and improved data processing proficiency, greatly enhancing the financial performance and competitiveness of mine enterprises.

Keywords: Comprehensive Budget Management (CBM); mine construction enterprises; integrated management information system; design and application

Yinzhou Zhu, Hui Yang, Baolin Yin. Method for Defining Multiple Homogeneous Activities in Distributed Workflow Management System

*Computer Modelling & New Technologies, 2014, **18**(1), 121-128*

In current process-oriented software systems, most of the processes have large number of parallel activities, which are homogeneous. These parallel activities are often used in the split-merge workflow structure and make the workflow model too complex to manage, as in the traditional workflow management systems each activity has to be defined respectively and bind to one resource. In this paper, we explore a novel method to define the distributed workflow model, which replaces the multiple homogeneous parallel activities with a batch-activity node to simplify the workflow model. An architecture is designed based on this method, which involves the model of organization structure, resource allocation and the sub-workflow. This architecture allows one batch-activity node bind to multiple resources, which are distributed, over a wide geographic area. Real-world scenarios, which are built and implemented based on this architecture, are shown to prove the effectiveness and usefulness of the method.

Keywords: workflow management, business process, distributed systems, resource allocation, multiple-instances pattern

Wang Mingpeng, Sun Qiming. Empirical Research on Existing Quantity of Small and Medium-sized Enterprises in China, Based on System Dynamics

*Computer Modelling & New Technologies, 2014, **18**(1), 129-135*

A system model for the existing quantity of small and medium-sized enterprises is built in this document, by establishing relationship equations with study on relationships among more than 30 variables such as total enterprise quantity, establishment rate of new enterprises, level of human resources, level of technical innovations, index of resource dependence, etc. Moderate breakthroughs are made on the mathematical methodology, such as the method of education years to calculate the level of human resources, the method of resource dependency evaluation for the resource dependency index. However, certain corrections are made for adaptation to the study. The innovative concept of establishment rate of new small and medium-sized enterprises is created in modelling and correlated with level of technical innovations, level of human resources and resource dependency index through relationship functions. The purpose thereof is to explore mechanisms where and extents to which influence factors make impact on the existing quantity of small and medium-sized enterprises. Finally, emulational prediction for the system model is made with the emulator Vensim and the error analysis on comparison between emulational and historical data is performed. It is found that the agreement with historical data is good and the error is acceptable.

Keywords: existing quantity of small and medium-sized enterprises, establishment rate of new small and medium-sized enterprises, system dynamics

Wang Fen, Chen Jianping. The Buyback Contract Coordination for a Logistics Service Supply Chain

*Computer Modelling & New Technologies, 2014, **18**(1), 136-143*

This article is about the coordination issue of the logistics service supply chain leading by the functional logistics provider (FLP). The service supply chain is consisted of the risk-neutral FLP and the loss-averse logistics integrator

(LI), and the contract model of the wholesale price and buyback contract model are established. The study found that the wholesale price contract cannot coordinate the supply chain, but the introduction of the buyback contract can stimulate the LI to increase the order quantity of the logistics capacity, reaching the level of the centralized logistics service supply chain and finally it is verified through examples.

Keywords: logistics service supply chain, loss averse, buyback contract, coordination

Bing Wang, Tinggui Chen, Guanglan Zhou, Chonghuan Xu. Research on resource-constrained project scheduling method based on heuristic priority rules

*Computer Modelling & New Technologies, 2014, **18**(1), 144-151*

The traditional project scheduling problem only considers the logical constraints. However, there is a need to consider the resources such as labour, capital and other constraints. Resource constrained project scheduling problem (RCPSP) integrates logic and resource constraints, which are closer to the actual scheduling. And most of these problems belong to NP-hard problem, which have certain difficulty in solving process. And therefore, there is a very important significance to study in the RCPSP, especially in the theory and application. Among the many the RCPSP models, resource constrained project scheduling problem is the most fundamental, but also the most basic model. Most RCPSP research has focused on a single project scheduling problem. The main content of this article describes the single resource constrained project scheduling problem and establishes the according model. Then, it studies how to use heuristic priority rules for solving the single project scheduling problem. In addition, this paper finally simulates a single project scheduling as an example of mold production and solves it by using the heuristic algorithm based on priority rules in order to verify the effectiveness of the algorithm. It combines with the different schedule generation schemes and priority rules as well as compares the different solution results. The final outcome indicates that the combination of different priority rules and schedule generation schemes would influence the single project scheduling results.

Keywords: resource constraints, project scheduling, heuristic priority rules

Xie Yuanmin. Optimization Research of Material Inventory Management Based on Genetic Algorithm

*Computer Modelling & New Technologies, 2014, **18**(1), 152-156*

Inventory management is an indispensable part in supply chain management. On the one hand the important position it has in the enterprise competitiveness. It has a direct impact on the cost of the products, respond speed to the market, delivery date and other indexes; On the other hand, the inventory system is a dynamic system, which involves huge inventory and wide area. The paper first under the influence of the uncertain factors does certain analysis to the raw material inventory issues, establishes minimum model of raw materials inventory cost in the iron and steel enterprises. Among them, the raw materials inventory cost includes fixed cost, procurement cost and storage cost. By making each purchase amount and raw materials inventory as the decision-making variables, it adopts the corrected genetic algorithm and uses MATLAB to get the optimal solution and get the most optimized raw materials procurement and inventory. The optimization model of raw materials inventory control, which presents in this paper mainly applies to iron and steel enterprises of continuous production process, and can also be extended to other types of the enterprises in raw materials inventory control.

Keywords: material inventory, iron and steel industry, optimization, genetic algorithm, MATLAB

Yehong Han, Lin Du. An update algorithm of decision rules in expert systems based on rough sets theory

*Computer Modelling & New Technologies, 2014 ,**18**(1), 157-164*

The decision table in rough sets theory is a kind of prescription, which specifies what actions should be undertaken when some of conditions are satisfied. Therefore, this tool can be used as knowledge representation system in expert systems. Decision rules, which are obtained by simplification of decision tables, can be used as rationale of decision reasoning. In order to compute new decision rules on the decision table in which a new instance is added, new instances are classified three cases according to the relation between the new instance and the original set of decision rules in the paper, and the category is proved that it is a partition of new instances. According to the category, an update algorithm of decision rules based on rough sets theory in expert systems is presented, and the complexity of the algorithm is obtained.

Keywords: rough sets, expert systems, incremental learning, decision table, algorithm

Iveta Kovalčíková, Juraj Kresila

Selected Contexts of Reform of National Education Programme in England and Slovakia

Kiev: Dragomanov National Pedagogical University – Institute of Ecology, Economy and Law
2012, 96 p, ISBN 978-966-660-812-6 (in Russian)


Ивета Ковальчикова, Юрай Кресила (2012) *Избранные контексты реформы национальной программы образования в Англии и Словакии.* Киев: Национальный педагогический университет имени М. П. Драгоманова – Институт экологии, экономики и права, 96 с. ISBN 978-966-660-812-6

The scientific research is devoted to the comparative analysis of the national education reform in England and Slovakia. The book is intended for scientific and pedagogical staff, graduate students, doctoral students, teachers, trainers, governmental education authorities, and other professionals dealing with education.

The main sections of the book:

- ✓ Contextual characteristics of the educational system.
- ✓ Curriculum as an educational factor.
- ✓ Comparison of education reforms and national curriculum development in the English and Slovak contexts.

The book has been prepared on the grounds of literary sources comprising historical and analytical essays revealing the development stages in the English school education. The focus of attention is on communications that directly preceded the adoption of the Reform Law [**Education Reform Act**] in 1988, and then on the analysis of some aspects of preparation and implementation of the education reform in England. Attention is also drawn to the arguments proving the legitimized need in education reform. The negative reasons of the contemporary state of things in education are considered in order to be identified and eliminated by the education reform. The general concept of the education reform is investigated and analyzed in detail. Particular attention is paid to the key issues of the National Curriculum development, to the general conceptual analysis of the major results in the development of school education in Slovakia, and to the comparative evaluation of the main characteristics in educational contexts of England and Slovakia.

About Authors

Iveta Kovalčíková **doc. Ph.Dr. Ph.D.**, Vice-rector of University of Prešov e-mail: iveta.kovalcikova@unipo.sk	She is professionally affiliated with the Department of Pre-school and Elementary Pedagogy and Psychology, Faculty of Education, University of Prešov, Slovak Republic. She is the national coordinator in the CiCe network - Children Identity and Citizenship Education in Europe (London Metropolitan University), former Fulbright research scholar at the Centre for the Psychology of Abilities, Competencies, and Expertise (PACE CENTER), Yale University, USA. Research focus: cross-cultural psychological research of minorities, dynamic assessment of cognitive abilities, compensatory education - focus on cognitive and executive functioning
Juraj Kresila **Ph.D.,** a lecturer at the Institute of British and American Studies, University of Prešov, Slovak Republic.	In his PhD dissertation, he analysed educational provisions of English and Slovak state curricula. His previous publications concerning the above topic include Curricular Reform in Slovakia in a Comparative Context, Journal of Pedagogy. Currently, his research orientation also involves cultural aspects of place semiotics.

Prof., Dr.Phil., Dr. Paed. **Tamara Lobanova-Shunina**