# CONTENTS

COMPREHENSIVE EVALUATION SYSTEM OF COLLEGE FOREIGN LANGUAGE CLOUD COMPUTING ASSISTED TEACHING..... XIAOHUA SHI 199

ARTIFICIAL BEE COLONY (ABC) OPTIMIZATION ALGORITHM FOR NONNEGATIVE LINEAR LEAST SQUARES.....XIANGYU KONG 202

APPLYING DATA MINING TECHNOLOGY TO SOLVE THE PROBLEM OF COACHING: A CASE STUDY..... JIN ZHANG 207

ANALYSIS OF MOTION OF ONE DIMENSIONAL DAMPING OSCILLATOR RESEARCH..... .....ZHOU SIHUA; YUAN HUANLI; YAO YAO; GUO YANHUA 211

APPLYING DATA MINING TECHNOLOGY TO SOLVE THE PROBLEM OF TRAFFIC: A CASE STUDY .....KE ZHU 214

TIME SENSITIVE PROXIMITY BASED ROUTING PROTOCOL IN OPPORTUNISTIC NETWORKS..... .....LI LIU 218

STUDY ON E-COMMERCE TRANSACTIONS CREDIT RATING OF CHINA'S FOREIGN TRADE ENTERPRISE..... LI HANWEI; LIU MENG 221

SOFTWARE APPLICATIONS IN DATA ACQUISITION AND ANALYSIS OF BATTERY PARAMETERS MEASUREMENT.....ZHANG LI-GUO, GAO JING 224

RESEARCH ON SPHERICAL PANORAMA MOSAIC TECHNOLOGY..... .....XUE LENG; SONG LI; JINHUA YANG 227

BPEL-BASED WEB SERVICE COMPOSITION AND CASE STUDY..... .....LIN LI; MING CHENG; YONG XU; GUOQING WU; AND MENGTING YUAN 231

EMPIRICAL STUDY OF DYNAMIC ENDOGENEITY: A PANEL VECTOR REGRESSION APPROACH..... .....LIN SHAO; BYUNGHUA LIM 234

THE WHEAT PRICES PREDICTION BASED ON ARIMA MODEL.....SHILIANG DU; DEXIAN ZHANG 239

FUZZY PROXIMAL SUPPORT VECTOR MACHINE FOR THE IMBALANCED AND BALANCED DATASETS.....CHUANDONG QIN 244

CALL FOR PAPERS.....248

# Comprehensive Evaluation System of College Foreign Language Cloud Computing Assisted Teaching

Xiaohua Shi
School of Foreign Language , Nantong University ,Jiangsu Nantong 226019

***Abstract*—In recent years, as the core technology of network computing platform, the intervention of cloud technology in the education field has expanded the developing space of foreign language educational technology and started the new thinking of the development of our domestic foreign language education career. This paper mainly discusses effective constructs of foreign language knowledge based on cloud learning environment and studies the comprehensive evaluation system of college foreign language cloud teaching deeply.**

***Index Terms*—Cloud computing, Could class, Independent exploration, Collaboration, Evaluation system**

## I. Introduction

In recent years, with the continuous deepening of college teaching reform, foreign language major in numerous colleges has been striving to attempt the deep reform of teaching mode so as to enrich the class teaching contents and promote the effective construct of students on professional knowledge of foreign languages. The intervention of cloud computing teaching assisted mode has greatly promoted the benign development of foreign language teaching and has become the effective supplement of traditional foreign language teaching. Although cloud technology teaching assisted mode owns numerous advantages, it still involves deficiencies. Therefore, we should establish the complete comprehensive evaluation system of foreign language cloud teaching so as to develop the function of cloud computing assisted foreign language teaching better.

## II. Theoretical Basis of the Introduction of Cloud Platform in College Foreign Language Major

Recognition teaching, as the idea of modern teaching, is the new teaching theory and practice produced for behaviorism teaching views and defects of exam-oriented teaching. Its essence is the emphasis on the student recognition development as the center so as to promote the development of student recognition to be teaching orientation.[1] Based on this teaching theory, there is a "bidirectional and interactive" communicative relationship between teachers and students.[2] And learning environment is the place learners study in "discovery style" and "exploration style". In this learning environment, students can utilize each learning tool and information resource to learn so as to reach the expected learning target.[3] The intervention of cloud computing in foreign language teaching can actually satisfy the characteristics of recognition teaching. The learning pattern based on cloud computing can promote the benign interaction between teachers and students. At the same time, students can utilize the mass foreign language resources provided by the cloud teaching platform to roam in the knowledge.

## III. The Effective Construct of Foreign Language Knowledge in the Cloud Learning Environment

### *A. The ingenious foreign language cloud class*

Traditional class focuses on the total quantity of taught knowledge, and students rarely have time to explore knowledge independently and think about problems. However, cloud class emphasizes on the knowledge standard of every student, the learning situation, how teachers implement individualized guidance and etc. Cloud learning platform can realize the co-collaboration, creation, edition and document share and etc, which provides very good interactive environment for learners. In the process of foreign language teaching, teachers and students are able to combine numerous application platforms in accordance with demands so as to establish the open and individualized cloud teaching environment.

1) Self-explored foreign language cloud class

Self-explored cloud class is the class teaching mode with the main characteristics of initiative exploration and the intensification of creative awareness, which pays more attention to the cultivation of the ability of learners in independent learning and provides the space of independent exploration learning in the creative education system so as to form the learning mode fitting the conditions for learners themselves. The intervention of cloud technology in foreign language teaching provides a good platform for learners and gives them abundant spaces of independent exploration. Students can access to cloud class only by inputting users and passwords in the cloud platform interface.

Differences of college foreign language major students are not obvious in the beginning period. However, after the learning of approximate two years, differences have become more and more outstanding and it has become difficult to cater for all widths and dimensions of class learning contents. Foreign cloud class fully emphasizes on individualized teaching, which

can satisfy the demand of implementing stratified teaching in the same class. Students can select their own learning modules in accordance with their learning ability situations and explore knowledge independently. Of course, teachers need to insist the principle that students are dominated and teachers are assisted.

2）Collaborated and shared foreign language cloud class

Cloud Computing Supported Collaboration Learning refers to the learning method using the technical method of cloud computing assisted teaching and supported collaboration learning, which builds the cloud environment in the clouds so as to allow each role of students and teachers to collaborate and communicate and is the extension and development of computer supported collaboration learning. [4] Foreign language teachers can base on cloud service, take collaboration learning as the principal line, match with each teaching assisted method and develop a series of collaboration learning activities in the clouds as the supplement of class teaching in expectation of cultivating the collaboration ability of students so as to improve the teaching quality.

### *B. Knowledge construct in the mobile learning mode*

Mobile learning, also referred to M-learning, is a kind of new learning mode that can be developed whenever and wherever possibly with the assistance of mobile computing device. It is the product of the combination of the three of wireless communication computing, Internet technology and modern education, equipped with characteristics of mobility, high efficiency, universality and individualization. [5] As a branch of digital learning, mobile learning has attracted more and more attention and gradually developed to be the new hot spot of foreign language education technology and related field research. As the necessary supplement of class teaching, for foreign language learners, the effective construct of foreign language knowledge is no long limited by space-time. In the cloud background, mass foreign language education resources are stored in the cloud server so that foreign language learners can use mobile devices whenever and wherever possibly (such as mobile phone, palm compute r(PDA), Pocket PC, Ipad and etc). With UC browser, we can connect to cloud storage, acquire mass foreign language learning resources and learn independently.

## IV. COMPREHENSIVE EVALUATION SYSTEM OF COLLEGE FOREIGN LANGUAGE CLOUD TEACHING

### *A. The necessities of establishing the complete teaching efficiency evaluation system*

As an important work of teaching management, basic link of teaching supervision and important move of teaching quality supervision, teaching evaluation has occupied a place in the entire teaching reform. Its essence is a kind of middle and microscopic education evaluation, which is also the process in which it investigates on the learning of students and the teaching of teachers systematically in accordance with certain teaching targets and standards and evaluates its values, strengths and defects so as to improve itself.[6] Traditional foreign language teaching evaluation emphasizes on the assessment of language knowledge, and modern foreign language teaching evaluation based on cloud technology needs to evaluate the teaching comprehensively in the general and developing perspectives, which needs to establish the complete teaching efficiency evaluation system.

### *B. The establishment of standardized evaluation system*

Teaching evaluation is a system project concerning on the long-term development of school and college construction, which has very strong guidance on the quality and level of teaching activity. Therefore, when improving the teaching evaluation system of teachers, it must fully consider the teaching contributions of vast teachers, respect the labors of numerous teachers sufficiently, evaluate the teaching level of teachers scientifically and improve the enthusiasm and initiatives of teachers in doing teaching activities.[7] Taking the college the author teaches in as the example, it adopts the four in one evaluation system of teacher self evaluation, student teaching assessment, teaching supervision, peer comment. In this way, it can make the objective and comprehensive comment on the foreign language teaching based on cloud computing so as to lay the solid foundation for the further improvement of foreign language cloud teaching.

### *C. Principles of cloud teaching evaluation*

Based on cloud computing technology, each foreign language teaching information can be filed to the united data base in real time so that the scientific teaching efficiency evaluation can be done by fully applying modern data analysis technology. As a complicated management activity, it must take certain principles as guidance if it wants to guarantee the quality of teaching evaluation and make evaluation results objective and scientific. Radically, the principle of teaching evaluation is the reflection of teaching essence and rules. At present, cloud teaching is still at the exploration stage, but the principles of exploration on the teaching evaluation is the basic requirements used to guide the work of exploring teaching evaluation work established by the people's understanding on exploring the teaching essence and rules.[8]

1) Crisscrossed principles

Vertical evaluation emphasizes on the developing trends and histories of evaluated objects, which is beneficial to exploring its developing rules. Horizontal evaluation should pay attention to the standard evaluated objects reach for horizontal comparisons. The both have advantages and disadvantages separately. Only by combining the both can it equip with scientificity and convenient for teaching evaluation to be perfect. In this research, vertical evaluation strives to explore the formation of cloud teaching idea of foreign language major, the allocation of cloud teaching hardware facility, the proficiency of teachers and students on the cloud platform and the sorting and exploration on the

developing trends and veins of foreign language teaching development. Horizontal evaluation stresses on the comparisons between foreign languages cloud teaching and traditional mode teaching and the teaching level it reaches at the current stage. The both supplement each other and none of them can be dispensed with. For foreign language in the cloud learning environment, it should insist the crisscrossed evaluation principle and make learners and teachers acquire improvements for constructing learning better so that it can actually promote the benign development of cloud teaching and implement the diversification of education information process.

2) Diversified principles

Diversified principles mean that themes, evaluation contents, evaluation standards and evaluation methods of education evaluation should be various instead of single.[9] Teaching evaluation in the traditional teaching mode depends on quantitative evaluation too much but it ignores qualitative evaluation, which usually makes evaluation cannot make the function of promoting teaching. For the evaluation of cloud teaching in applications on modern education technology, it should also insist the diversification, emphasize the participation and interaction, combine comments on ourselves and others, melt formative comments and summary comments, strive to equip with objectivity and scientificity and implement the promotion of learning with comments.

3) Effective principles

Effective principles can be said that education evaluation process should draft practical evaluation plans, adopt proper evaluation methods and make evaluations provide guidance and assistance to education activities effectively in accordance with goals and contents of evaluation, specific situations of evaluation objects, specific social environment backgrounds and etc. [10] It should be said that effective principles run through the teaching process of each foreign language major and each link of the foreign language teaching process. It not only has important meanings in the aspect of professional knowledge teaching in foreign language major, but also develops positive functions in the educational aspect of other non-cognitive factors. Based on this principle, it mainly investigates the feasibility of cloud computing technology in the practice of foreign language teaching in the evaluation process as well as the liabilities of good results produced by cloud teaching so as to provide solid evidences for teaching evaluation. It is only by insisting in effective principles that it can fall the purpose of improving student foreign language comprehensive ability into place.

By constructing the basic frame mode of college foreign language cloud teaching comprehensive evaluation system (shown in Fig. 1), it can promote the constant improvement of cloud computing assisted foreign language teaching level effectively.

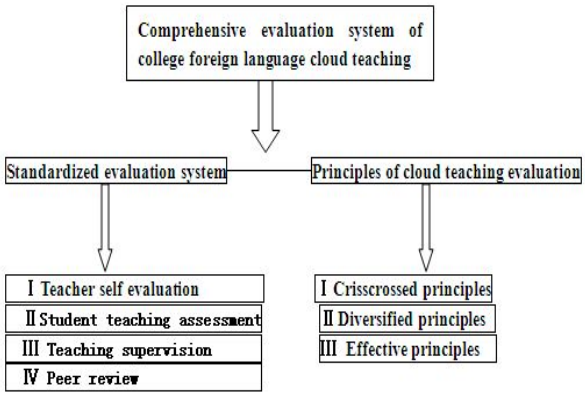



Figure.1 basic frame mode of college foreign language cloud teaching comprehensive evaluation system

## V. CONCLUSION

Cloud computing assisted teaching mode injected new vigor into the traditional foreign language teaching, promoting the benign development of foreign language education effectively. By constructing the comprehensive evaluation system of foreign language cloud teaching, it can regulate foreign language teaching scientifically, promote it to develop its positive sides and make it become the necessary supplement of foreign language teaching.

## REFERENCE

[1] Shijin Sun. Introduction of Psychology. Shanghai: Huadong Normal University Press, 2002: 161.

[2] Weixin Yuan. Test Theory on Scientific Education Ideas based on Constructionism. Foreign Education Research, 2002(6): 3-4.

[3] Jianxin Zhu. English-Chinese Interpretation Network Teaching Mode under the View of Recognition Teaching Theory. Foreign Language Audiovisual Education,2009(9): 61.

[4] Li Zhou.Cloud Computing Supported Collaboration Learning: Take Education Technical Principles Course as the example. Shanghai Normal University,2010: 3.

[5] Jinguo Mo.Foreign Language E-learning Theory and Practice. Shanghai: Shanghai Foreign Language Education Press, 2011: 41.

[6] Xiaorong Li, Longyu Wei. Education Evaluation. Chengdu, Sichuan Education Press: 1988: 4.

[7] Lei Du, Yan Zhang, Youfei Guan. Proposals of Teaching Evaluation System for Teachers in Research-based Colleges. Academic Journal of Shanxi Medical University (Basic Medical Educational Edition), 2009(4): 501-503.

# Artificial Bee Colony (ABC) Optimization Algorithm for Nonnegative Linear Least Squares

Xiangyu Kong
Institute of Mathematics and Information Science, Beifang University of Nationalities, Yinchuan, China
Email: kxywz08@163.com

***Abstract*—This paper presents the comparison results on the performance of the Artificial Bee Colony (ABC) algorithm for Nonnegative Linear Least Squares. The ABC algorithm has been firstly proposed for unconstrained optimization problems and showed that it has superior performance on these kind of problems. In this paper, the ABC algorithm has been extended for solving NLLS problems. Numerical results are reported which demonstrate very good computational performance.**

***Index Terms*—Artificial Bee Colony algorithm, Nonnegative Linear Least Squares, Optimization Problem**

## I. Introduction

The method of least squares is a standard approach to the approximate solution of over determined systems, i.e. sets of equations in which there are more equations than unknowns [1]. Least squares problems fall into two categories: linear least squares and non-linear least squares, depending on whether or not the residuals are linear in all unknowns [2]. In statistics and mathematics, linear least squares is an approach to fitting a mathematical or statistical model to data in cases where the idealized value provided by the model for any data point is expressed linearly in terms of the unknown parameters of the model. The resulting fitted model can be used to summarize the data, to predict unobserved values from the same system, and to understand the mechanisms that may underlie the system [3].

The Nonnegative Linear Least Squares (NLLS) problem [4],

$$\min_{x\geq 0} f(x)=\frac{1}{2}\|Ax-b\|^2=\frac{1}{2}(Ax-b)^T(Ax-b) \quad (1)$$

Where $A\in R^{m\times n}$ with $m\geq n$, $rank(A)=n$, and $b\in R^m$, is a common optimization problem arising from many diverse applications.

Over the past few years, a number of approaches have been developed for solving the linear least squares using classical mathematical programming methods [4-5]. These methods require matrix factorizations or updates, and can become overly expensive for very large-scale problems.

Gradient-type methods, such as gradient projection methods [6], require matrix-vector multiplications, but typically have very slow convergence. Meanwhile, classical optimization methods are highly sensitive to starting points and frequently converge to local optimum solution or diverge altogether.

Karaboga has described an Artificial Bee Colony (ABC) algorithm based on the foraging behavior of honey bees for numerical optimization problems [7-8].Karaboga and Basturk have compared the performance of the ABC algorithm with those of other well-known modern heuristic algorithms such as Genetic Algorithm (GA), Differential Evolution (DE), and Particle Swarm Optimization (PSO) on unconstrained problems. In this work, ABC algorithm is extended for solving Nonnegative Linear Least Squares problems.

Paper is organized as follows. In Section 2, the ABC algorithm is introduced. Simulation results of ABC, HS[9], HSCH[10] and HSWB[11] are presented and compared in Section 3.Finally, a conclusion is provided.

## II. Artificial Bee Colony Algorithm for Nonnegative Linear Least Squares Problems

In ABC algorithm, the colony of artificial bees consists of three groups of bees: employed bees, onlookers and scouts. First half of the colony consists of the employed artificial bees and the second half includes the onlookers. For every food source, there is only one employed bee. In other words, the number of employed bees is equal to the number of food sources around the hive. The employed bee whose the food source has been abandoned by the bees becomes a scout.

In ABC algorithm, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. At the first step, the ABC generates a randomly distributed initial population $P(G=0)$ of $SN$ solutions (food source positions), where $SN$ denotes the size of population. Each solution $x_i(i=1,2,...,n)$ is a $D$-dimensional vector. Here, $D$ is the number of optimization parameters. After initialization, the population of the positions (solutions) is subjected to repeated cycles, $C=1,2,...,MCN$, of the search processes of the employed bees, the onlooker bees and scout bees. An employed bee produces a modification on the position (solution) in her memory depending on the local information (visual information) and tests the nectar amount (fitness value) of the new source (new solution). Provided that the nectar amount of the new one is higher than that of the previous one, the bee memorizes the new position and forgets the old one.

Otherwise she keeps the position of the previous one in her memory. After all employed bees complete the search process, they share the nectar information of the food sources and their position information with the onlooker bees on the dance area. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount. As in the case of the employed bee, she produces a modification on the position in her memory and checks the nectar amount of the candidate source. Providing that its nectar is higher than that of the previous one, the bee memorizes the new position and forgets the old one.

## III. COMPUTATIONAL RESULTS AND COMPARISONS

In this section we perform some NLLS problems in order to illustrate the implementation and efficiency of the ABC method.

NLLS 1. Consider the following NLLS problem, where

$$A=\begin{bmatrix}1 & 1 & 1\\ 0 & 4 & -1\\ 2 & -2 & 1\end{bmatrix},\ b=\begin{pmatrix}6\\ 5\\ 1\end{pmatrix}$$

The optimal solution is $x^*=(1,1,3)^T$.

NLLS 2. Let be a matrix whose diagonal elements are 500 and the nondiagonal elements are chosen randomly from the interval such that $A$ is symmetric. Let $b=Ae$ where e is $n\times 1$ vector whose elements are all equal to unity such that $x^*=(1,1,...,1)^T\in R^n$ is the unique solution.

NLLS 3. Let the matrix $A$ is given by $a_{i,i}=4n$, $a_{i,i+1}=a_{i+1,i}=n$, $a_{i,j}=0$, $i=1,2,...,n$. Let $b=Ae$. Thus the unique solution is $x^*=(1,1,...,1)^T\in R^n$.

NLLS 4. Following we consider one randomly generated NLLS problem where the data $(A,b)$ are generated by the Matlab scripts: $rand('state',0)$; $m=200$; $n=100$; $A=rand(m,n)$; $b=A*ones(n,1)$, where $A\in R^{m\times n}$ ,and the unique solution is $x^*=(1,1,...,1)^T\in R^n$.

NLLS 5. Following we consider another randomly generated NLLS problem where the data $(A,b)$ are generated by the Matlab scripts: $rand('state',0)$; $n=100$; $A1=rand(n,n)$; $A=A1'*A1+n*eye(n,n)$; $b=A*ones(n,1)$, here $A$ is a positive definite symmetric matrix, and the unique solution is $x^*=(1,1,...,1)^T$.

To judge the accuracy of different algorithms, 30 independent runs of each of the four algorithms were carried out and the best, the mean, the worst fitness values, and the standard deviation (Std) were recorded. Table 1 compares the algorithms on the quality of the optimum solution for given NLLS problems.

TABLE 1: THE STATISTICAL RESULTS FOR 30 RUNS TESTED ON GIVEN NLLS PROBLEMS.

| Functions | Algorithms | Best | Mean | Worst | Std |
|---|---|---|---|---|---|
| NLLS1 | HS | 1.05e-07 | 1.06e-02 | 9.22e-02 | 1.92e-02 |
| | HSCH | 6.30e-05 | 8.39e-02 | 6.05e-01 | 1.20e-01 |
| | HSWB | 2.11e-06 | 5.88e-02 | 4.57e-01 | 9.85e-02 |
| | ABC | 1.39e-17 | 7.58e-17 | 2.77e-16 | 5.92e-17 |
| NLLS 2 | HS | 7.64e+07 | 8.84e+07 | 9.94e+07 | 5.80e+06 |
| | HSCH | 3.21e+06 | 4.17e+06 | 5.33e+06 | 4.96e+05 |
| | HSWB | 4.61e+06 | 6.29e+06 | 9.39e+06 | 1.05e+06 |
| | ABC | 5.52e-11 | 4.61e-10 | 1.62e-09 | 3.36e-10 |
| NLLS 3 | HS | 5.12e+07 | 6.79e+07 | 7.80e+07 | 5.96e+06 |
| | HSCH | 2.00e+06 | 2.69e+06 | 3.33e+06 | 2.80e+05 |
| | HSWB | 2.77e+06 | 3.85e+06 | 5.56e+06 | 7.06e+05 |
| | ABC | 9.31e-03 | 2.06e-02 | 4.92e-02 | 8.25e-03 |
| NLLS 4 | HS | 2.93e+05 | 3.52e+05 | 4.53e+05 | 4.20e+04 |
| | HSCH | 4.02e+02 | 5.45e+02 | 6.91e+02 | 6.77e+01 |
| | HSWB | 2.96e+02 | 4.11e+02 | 5.14e+02 | 5.63e+01 |
| | ABC | 2.72e+00 | 9.27e+00 | 1.93e+01 | 4.15e+00 |

| | | | | | |
|---|---|---|---|---|---|
| NLLS 5 | HS | 3.12e+08 | 4.71e+08 | 6.10e+08 | 5.89e+07 |
| | HSCH | 4.17e+05 | 5.22e+05 | 6.29e+05 | 4.97e+04 |
| | HSWB | 2.29e+05 | 3.29e+05 | 4.03e+05 | 4.50e+04 |
| | ABC | 2.96e+02 | 1.17e+03 | 5.35e+03 | 1.22e+03 |

Figs. 1-5 show the convergence and its boxplot figure of the best fitness in the population for the different algorithms. The values plotted for every generation are averaged over 30 independent runs. As can be seen, the ABC algorithm is the best not only for simple NLLS problems, but also for complex NLLS problems.

## IV. CONCLUSION

We have given ABC algorithm for solving the NLLS problem. The ABC has strong global search ability in the early stage of optimization, and has strong local search ability in the late stage of optimization. This ensures that the explorative power of ABC is on average greater than that of the other three methods, which in turn results into better accuracy of the ABC algorithm.

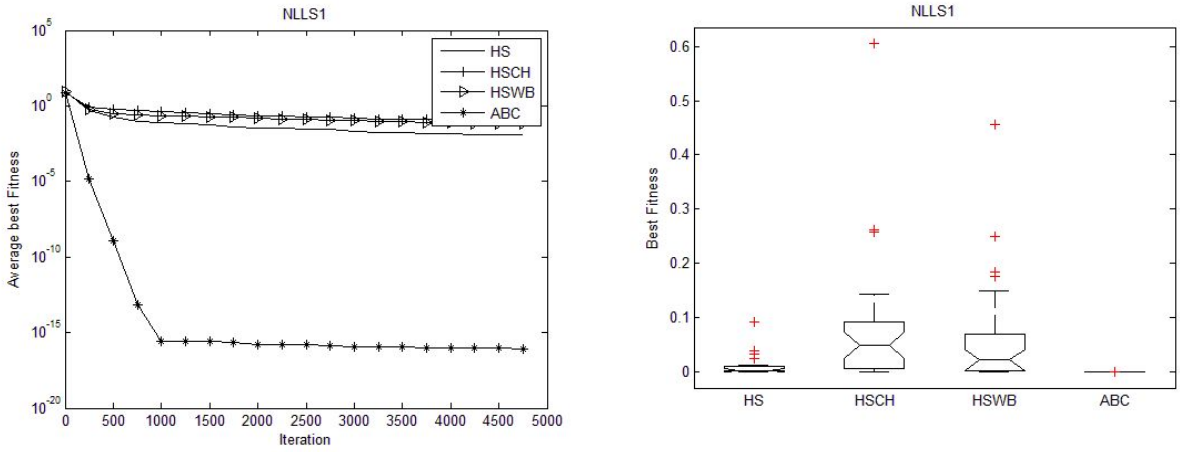



Figure1. The convergence and its boxplot of the best fitness for NLLS1 with $n = 3$ .

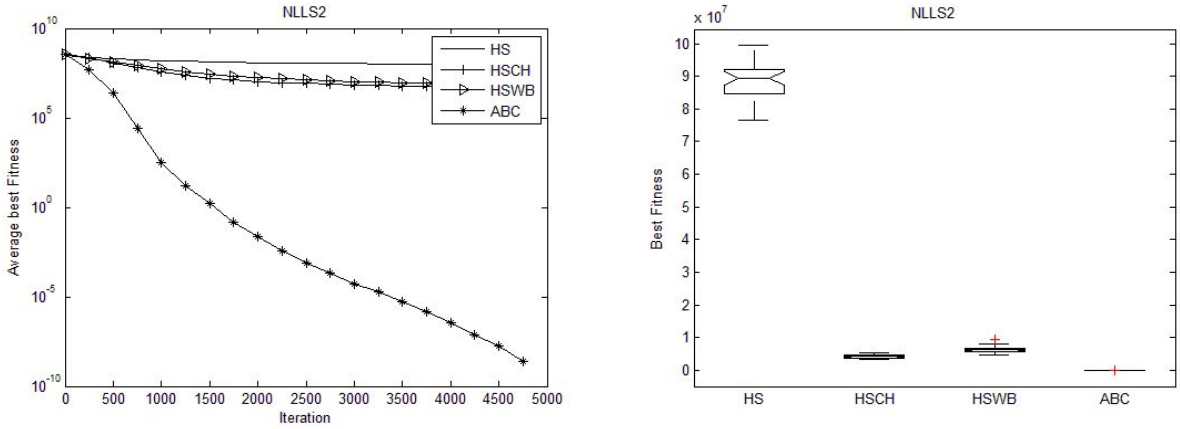



Figure2. The convergence and its boxplot of the best fitness for NLLS2 with $n = 100$ .

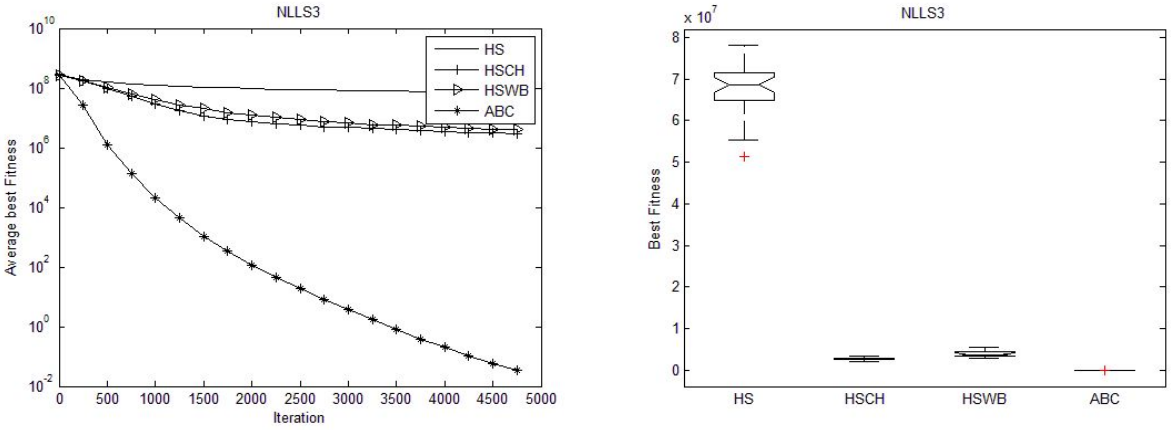



Figure3. The convergence and its boxplot of the best fitness for NLLS3 with $n = 100$ .

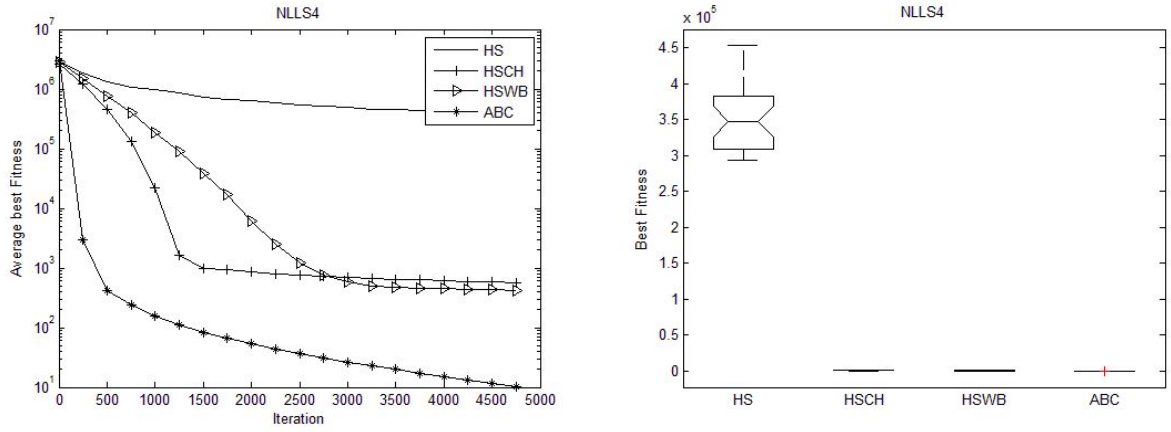



Figure4. The convergence and its boxplot of the best fitness for NLLS4 with $n = 100$ .

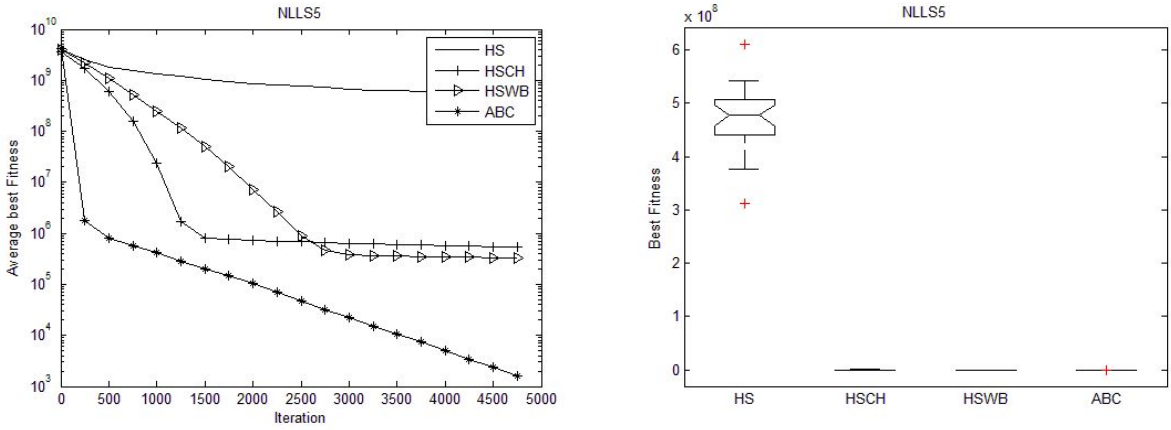



Figure5. The convergence and its boxplot of the best fitness for NLLS5 with $n = 100$ .

## ACKNOWLEDGMENT

This work is supported by 2013 Narure Science Foundation of Ningxia (No. NZ13096), 2013 Higher educational scientific research project of Ningxia (No. NGY2013086), 2013 scientific research project of Beifang University of Nationalities (2013XYZ021,2014XBZ01).

## REFERENCES

[1] J. Nocedal, S. J. Wright, Numerical Optimization, Springer, 1999

[2] C. L. Lawson, R. J. Hanson, Solving least squares problems, Society for Industrial and Applied Mathematics, 1974, 9-256

[3] Bjorck, Ake, Numerical methods for least squares problems, Philadelphia: SIAM, 1996

[4] Longquan Yong, A feasible interior point algorithm for a class of nonnegative least squares problems,ICIC Express Letters, 4 (2010), 131-136
[5] Longquan Yong, Sanyang Liu, Jianke Zhang et al., An improved harmony search algorithm for linear least squares, Journal of Computational Information Systems, 8 (2012), 4217-4224
[6] G. P. McCormick, R. A. Tapia, The gradient projection method under mild differentiability conditions,SIAM J. Control, 10 (1972), 93-98
[7] D. Karaboga, An Idea Based On Honey Bee Swarm For Numerical Optimization, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
[8] B. Basturk, D.Karaboga, An Artificial Bee Colony (ABC) Algorithm for Numeric function Optimization, IEEE Swarm Intelligence Symposium 2006, May 12-14, 2006, Indianapolis, Indiana, USA.
[9] Z. W. Geem, J. H. Kim, G. V. Loganathan, A new heuristic optimization algorithm: Harmony search, Simulation, 76 (2001), 60-68
[10] Shouheng Tuo, Longquan Yong, Improved harmony search algorithm with chaos, Journal of Computational Information Systems, 8 (2012), 4269-4276

Longquan Yong et al., Harmony Search Based on Worst-Best Strategy for Absolute Value Equation, Technical Report, School of Mathematics and Computer Science, Shaanxi University of Technology, China, 2012

# Applying Data Mining Technology to Solve the Problem of Coaching: A Case Study

Jin Zhang
Department of Educational Technology, Henan Normal University, Xinxiang, China
Email:zj@htu.cn

***Abstract*—In order to find the best all time college coach in last century, this paper build two models. In model one, we introduce a simple method of evaluating coach through analytic hierarchy process (AHP). In model two, we have obtained the specific data from the authentic website and have a series of processing to it through the given indicators so that we can get the ranking of college coaches of the sports that the problem demands, finally, we find the best college coach. Then, we analyze the effect of different time line, which also means the different processing of evaluation indicators in different time line to make our models have a wider range of application. Meanwhile, we make some modifications so that it can be applied to both genders and all possible areas. So, our models have reached a satisfactory degree. At the request of problem, we choose three sports and list the top five coaches in each sport. We also point out the strengths and weaknesses of our models and some Prospects thinking for the future study.**

***Index Terms*—Analytical hierarchy process, Statistical Analysis, Fuzzy comprehensive analysis**

## I. INTRODUCTION

Sports Illustrated is an American sports media franchise owned by media conglomerate Time Warner. The scouting reports—including a World Series Preview and New Year's Day bowl game round-up that enhanced the viewing of games on television. Now, Sports Illustrated holds an activity to find the people called the best all time coach in college sports in the history of the last century. This paper use mathematical models to make solutions to find the best all time college coach in the history of the last century and list the top five college sports coaches in three different sports.

Coach is the most important figure in a sports team .We can say that whether a coach is good or not will basically decide whether a team will win or not. There were many excellent coaches in the last century [1]. But who would be the best? Are there any indicators that can judge a coach is excellent or not? What's more, does it make a different in time line horizon, gender different or sports? This paper will solve all these problems. Besides, according to the requirements of the problem, we will study such sports as college hockey, football, basketball, baseball to choose the best college coach for the previous century in this paper.

A brilliant college coach must qualify himself many excellent abilities of various aspects such as the team's winning rate, the time of teaching and so on. We must find out and analyze these elements, measure their importance, give the weights and finally get the formula to calculate the overall performance score that represent coaches' excellence degree. According to the final analysis of the overall performance score, we can know whether a coach is good or not and his ranking.

During the previous century's development of college P.E, there were several important time line horizons. For example, after the World War I, people gradually attached the importance to college physical culture, and the need of college coach began to expand. As to the World War II, the college physical culture developed faster and the university sports coaches' demand was in short supply. In the 1960s, the needs of college tended to balance [2].

When it comes to 20th century while feminism was on the wane, people thought that women couldn't organize and train a team like men especially in baseball and football where women were in obvious inferior strength. So the rate of women college coaches is very low in previous century. Today however, feminism caught the attention of people and the rate began to be higher.

## II. METHODOLOGY

### A. Problem Analyses

According to our life experiences and the materials we've collected, we've found out a serious of indicators to measure whether a coach is good or not, when we work out all the weights of all the indicators through the mathematical model, we can build a function based in all the indicators. And then, we can take all the indicators that we've collected into the function to get the final score. It will be the final grade of each coach which represents the level of a coach's excellence. Finally, by comparing the scores, we can get the ranks and pick up the best coach in previous century.

From the problem, we can know the mainly elements which influence the model's universal serviceability.

(1) The time axis: As the college's development was different in different time line in the previous century, the indicators to measure the coach's excellence can change slightly. But it's hard to control the change. If the indicators we use keep unchanging, it will be unfair for some coaches.

(2)The coaches' gender: Now that there are male coaches as well as female coaches, the indicators should change slightly.

(3)The sports: Since the caches taught different sports, the indictors should change slightly.

According to the requirements of the problem, we should pick up 3 sports first and then found the top 5 coaches in each different sports according to the model. Actually, this requirement's aimed at testing the model. In other words, we should compare the 15 coaches we find with the real data to see the degree of compliance and guarantee its reliability.

*B. Assumptions*

We make some assumptions about the problem we solved in this paper:

- We assume the coach's gender does not affect the evaluation indicators in this model for the time being.
- We assume the range of time does not affect the evaluation indexes in this model for the time being.
- We assume the sports do not affect the evaluation indexes in this model for the time being.

*C. Analysis*

We choose four sports: baseball, football, and basketball and college hockey in this model.

In football for example, first, we look for the specific information of the most distinguished 10 college football coaches including the number of games that they guided, the number of games that the team they guided wins, losses, and sites, and the probability of win. Then, we can get the level of the coaching skills of the coaches through sort their games by some certain indicators. Last, we get the rank of the 10 coaches by the fuzzy comprehensive evaluation to their works.

We can also use this method to get the rank of the 10 coaches by the fuzzy comprehensive evaluation to their works at baseball, basketball and soccer.

About the background of the fuzzy comprehensive evaluation, we know the professor L.A.Zadeh created the fuzzy mathematics at 1965, at the same time the application of it in the two aspects of theory and practical has achieved great results [3].

- Pct.: Win-Loss percentage
- R: The evaluation matrixes of 10 coaches.
- W: The weights set of the four indicators {Win-Loss, Pct, Years}.
- S: The fuzzy judgment set.
- μ: The evaluation result.

We make the evaluation factors set is {Win-Loss, Pct, Years}, and the weights set is W= {0.5, 0.3, 0.2}.

The evaluation matrixes of 10 coaches are:

$$R_1 = \begin{bmatrix} 0.6651 & 0.0004 & 0.3345 \\ 0.0000 & 0.3074 & 0.6926 \\ 0.0000 & 0.8700 & 0.1300 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.6805 & 0.0033 & 0.3162 \\ 0.0000 & 0.4815 & 0.5185 \\ 0.0000 & 0.4815 & 0.5185 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0.7306 & 0.0012 & 0.2682 \\ 0.0000 & 1.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} 0.7422 & 0.0017 & 0.2561 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \end{bmatrix}$$

$$R_5 = \begin{bmatrix} 0.7979 & 0.0009 & 0.2012 \\ 0.0000 & 1.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

$$R_6 = \begin{bmatrix} 0.6357 & 0.0017 & 0.3626 \\ 0.0000 & 0.0000 & 1.0000 \\ 0.0000 & 0.5460 & 0.4540 \end{bmatrix}$$

$$R_7 = \begin{bmatrix} 0.6548 & 0.0031 & 0.3421 \\ 0.0000 & 0.0000 & 1.0000 \\ 1.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

$$R_8 = \begin{bmatrix} 0.6367 & 0.0035 & 0.3598 \\ 0.0000 & 0.0000 & 1.0000 \\ 0.0000 & 0.6878 & 0.3122 \end{bmatrix}$$

$$R_9 = \begin{bmatrix} 0.8083 & 0.0073 & 0.1844 \\ 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

$$R_{10} = \begin{bmatrix} 0.7919 & 0.0011 & 0.2070 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \end{bmatrix}$$

We can take the operator and make sure the fuzzy judgment set S, then we can judge by the principle of maximum membership degree:

$$S_1 = W \circ R_1 = (S_K)_{1*N} = (0.5\ 0.3\ 0.2) \circ \begin{bmatrix} 0.6651 & 0.0004 & 0.3345 \\ 0.0000 & 0.3074 & 0.6926 \\ 0.0000 & 0.8700 & 0.1300 \end{bmatrix} = (0.5\ 0.3\ 0.3345)$$

$$S_2 = W \circ R_2 = (S_K)_{1*N} = (0.5\ 0.3\ 0.2) \circ \begin{bmatrix} 0.6805 & 0.0033 & 0.3162 \\ 0.0000 & 0.4815 & 0.5185 \\ 0.0000 & 0.4815 & 0.5185 \end{bmatrix} = (0.5\ 0.3\ 0.3162)$$

$$S_3 = W \circ R_3 = (S_K)_{1*N} = (0.5\ 0.3\ 0.2) \circ \begin{bmatrix} 0.7306 & 0.0012 & 0.2682 \\ 0.0000 & 1.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \end{bmatrix} = (0.5\ 0.3\ 0.2682)$$

$$S_4=W\circ R_4=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.7422 & 0.0017 & 0.2561\\ 0.0000 & 1.0000 & 0.0000\\ 0.0000 & 1.0000 & 0.0000\end{bmatrix}=(0.5\ 0.3\ 0.2561)$$

$$S_5=W\circ R_5=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.7979 & 0.0009 & 0.2012\\ 0.0000 & 1.0000 & 0.0000\\ 1.0000 & 0.0000 & 0.0000\end{bmatrix}=(0.5\ 0.3\ 0.2012)$$

$$S_6=W\circ R_6=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.6357 & 0.0017 & 0.3626\\ 0.0000 & 0.0000 & 1.0000\\ 0.0000 & 0.5460 & 0.4540\end{bmatrix}=(0.5\ 0.2\ 0.3626)$$

$$S_7=W\circ R_7=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.6548 & 0.0031 & 0.3421\\ 0.0000 & 0.0000 & 1.0000\\ 1.0000 & 0.0000 & 0.0000\end{bmatrix}=(0.5\ 0.0031\ 0.3421)$$

$$S_8=W\circ R_8=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.6367 & 0.0035 & 0.3598\\ 0.0000 & 0.0000 & 1.0000\\ 0.0000 & 0.6878 & 0.3122\end{bmatrix}=(0.5\ 0.2\ 0.3598)$$

$$S_9=W\circ R_9=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.8083 & 0.0073 & 0.1844\\ 1.0000 & 0.0000 & 0.0000\\ 1.0000 & 0.0000 & 0.0000\end{bmatrix}=(0.5\ 0.0073\ 0.1844)$$

$$S_{10}=W\circ R_{10}=(S_K)_{1*N}=(0.5\ 0.3\ 0.2)\circ\begin{bmatrix}0.7919 & 0.0011 & 0.2070\\ 0.0000 & 1.0000 & 0.0000\\ 0.0000 & 1.0000 & 0.0000\end{bmatrix}=(0.5\ 0.3\ 0.2070)$$

If 70% of the coaches that ranked by the analytic hierarchy process (AHP) are the latter half of last century, we can take the way two. Otherwise, we can take the way one.

The way one: if the effect isn't too big, we can mention the coaching time in the indicators of evaluation and consider it, we can do some adjustment to the weight by the characteristics of different sports.

The way two: if the effect is too big or can't be controlled, we should do the overall adjustment to the coaching time on the result. The adjustment is: dividing the whole century to different extents respectively and divided them by the corresponding proportion (such as the 60s-80s, we can make the result divided by 0.85, but after the 2000, we can make the result divided by 0.95).

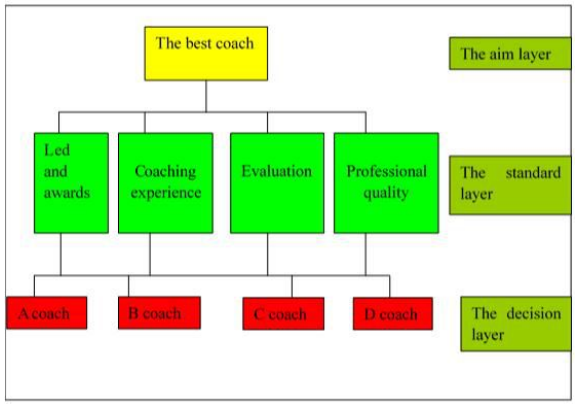



Figure 1. An Analysis Modeling

Table I. the level of the coaching skills of the coaches

| Coaches | Wins | Losses | Pct | Years |
|---|---|---|---|---|
| Mike Krzyzewski | 975 | 302 | .764 | 38 |
| | 73 | 59 | .553 | 5 |
| | 902 | 243 | .788 | 33 |
| Harry Statham | 1076 | 444 | .708 | 47 |
| Danny Miles | 1000 | 409 | .710 | 42 |
| Herb Magee | 976 | 391 | .714 | 45 |
| Jim Boeheim | 942 | 314 | .750 | 37 |
| Don Meyer | 923 | 324 | .740 | 36 |
| | 37 | 41 | .474 | 3 |
| | 665 | 179 | .788 | 23 |
| | 221 | 104 | .680 | 10 |
| Bob Knight | 902 | 371 | .709 | 42 |
| | 102 | 50 | .671 | 5 |
| | 662 | 239 | .735 | 30 |
| | 138 | 82 | .627 | 7 |
| Dean Smith | 879 | 254 | .776 | 36 |
| Adolph Rupp | 876 | 190 | .822 | 41 |
| Jim Calhoun | 873 | 380 | .701 | 39 |
| | 248 | 137 | .644 | 14 |
| | 625 | 243 | .726 | 25 |

If 70% of the coaches that ranked by the analytic hierarchy process (AHP) are the latter half of last century, we can take the way two. Otherwise, we can take the way one.

The way one: if the effect isn't too big, we can mention the coaching time in the indicators of evaluation and consider it, we can do some adjustment to the weight by the characteristics of different sports.

The way two: if the effect is too big or can't be controlled, we should do the overall adjustment to the coaching time on the result. The adjustment is: dividing the whole century to different extents respectively and divided them by the corresponding proportion (such as the 60s-80s, we can make the result divided by 0.85, but after the 2000, we can make the result divided by 0.95).

The traditional evaluation standard to the college coaches is too single, and the level of coaches in different years can't be regarded as the same. The result of the evaluation to coaches are affected by the professional quality of the coach, sports equipment, the popular of sports culture, the support comes from schools and institutions and other factors. So, we should regard this factor as the indicator of the evaluation to coaches.

## III. Conclusions and Further Explanations

We solved four problems in this paper:

1. We find the best all time college coach in last century through establishing mathematical model.

2. We analyze the effect of different time line horizon and eliminate the effect try our best.

3. We discuss how the model can be extended to both genders and all possible sport fields.

4. We choose three different sports and show the top five coaches of each sport.

The results:

(1)We establish the two models by respectively using the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation, and we choose the best all time college coach in last century is a basketball coach whose name is Adolph Rupp through calculate the certain data in the model.

(2)We explain the different time line horizons have effect to the result, and we give the solution through the analytic hierarchy process (AHP).

(3)We discuss how the model can be extended to both genders and all possible sport fields, and then we modified the indicators of the model and get the solution through the fuzzy comprehensive evaluation method.

(4)We choose three different sports and show the top five coaches of the model. The followed table is the result.

## Acknowledgment

This work is supported by soft-scientific project of Henan Province (No. 142400410425)

## References

[1] http://www.statesman.com/news/sports/top-college-football-coaches-of-all-time/nSztQ/

[2] Nielsen, J. (2000). Designing web usability: The practice of simplicity. Indianapolis, IN: New Riders Publishing.

[3] Spielberger, C. D. (1976). The nature and measurement of anxiety. In C. S. Spielberger (Ed.), Cross-cultural anxiety. Washington, DC: Hemisphere/Wiley.

**Zhang Jin**: born in 1983, Her research interests include Computer education.

# Analysis of Motion of One Dimensional Damping Oscillator Research

Zhou Sihua; Yuan Huanli; Yao Yao; Guo Yanhua
School of Physics and Electromechnical Engineering, Zhoukou Normal University, Henan, China

***Abstract*: In order to analyze the forced vibration of damping oscillator in physical problems, the different forms of Duffing equation have been studied in detail. The solutions of equation have been simulated with Matlab, and the vibration period and the frequency variation have been analyzed and compared. The forced vibration term, nonlinear term and damping term have different contribution to the solution of oscillator motion. The simulation results are in accordance with the current conclusions, this is an intuitive way to understand problems.**

***Index Terms*: forced vibration; Duffing equation; Matlab; nonlinear**

## I. INTRODUCTION

Spring oscillator is a typical model of the vibration, of which the motion states significantly change according to the different external conditions, mainly consist of the forced vibration term, nonlinear term and damping term. In traditional instruction, only the model of harmonic vibration is discussed, while the forced vibration and nonlinear conditions are not instructed. Fortunately, the physical phenomena described by Duffing equation[1] coincide with spring oscillator model also exists in the electromagnetic field.

## II. ANALYSIS AND SIMULATION PROCESS

We often encounter the questions of solving the forced vibration of the classical harmonic oscillator with unit mass and damping in the textbooks, where the questions boil down to linear and nonlinear questions[2-7]. Firstly, the simpler linear question is discussed, and the differential equation of harmonic oscillator is

$$\ddot{x}(t)+2\gamma\dot{x}(t)+\omega_0^2 x(t)=f(t) \tag{1}$$

Where f(t) is a known function, and we take e it as $\cos\Omega t$ for ease of calculation; $\gamma$ and $\omega_0$ are constant, which can be seen in the third edition of the method of mathematical physics edited by Yao Duanzheng, page 168[8]. Equation (1) is one form of Duffing equation.

$$\ddot{x}(t)+2\gamma\dot{x}(t)+\omega_0^2 x(t)+\varepsilon\beta_0^2 x^3=f(t) \tag{2}$$

if $\gamma,\varepsilon,f(t)$ in *Eq.*(2) are all zeros, which means removing the damping force, the nonlinear restoring force and external force, then expression turnes into the harmonic vibration equation

$$\ddot{x}(t)+\omega_0^2 x(t)=0 \tag{3}$$

It is easy to get

$$x(t)=A\cos(\omega_0 t+\theta) \tag{4}$$

The result of simulating Eq (4) is shown in Fig.1.

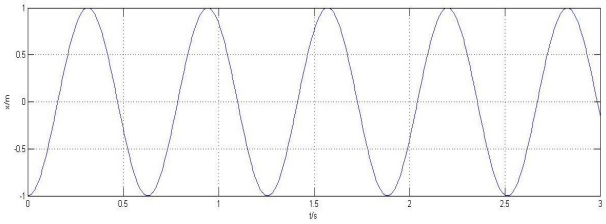


Figure.1 Simple harmonic vibration image of one-dimensional harmonic oscillator

In the textbook Eq.(1) is Fourther transformed with the Fourier transform method, and the ordinary differential equation has been obtained.

$$(i\omega)^2\tilde{x}(\omega)+2\gamma(i\omega)\tilde{x}(\omega)+\omega_0^2\tilde{x}(\omega)=\tilde{f}(\omega) \tag{5}$$

The result is

$$\tilde{x}(\omega)=\frac{-\tilde{f}(\omega)}{\omega^2-i2\gamma\omega-\omega_0^2}=\frac{\tilde{f}(\omega)}{(\sqrt{\omega_0^2-\gamma^2})^2-(\omega-i\gamma)^2} \tag{6}$$

Using the transform properties and by table lookup, the solution can be written as

$$x(t)=c_0 e^{-\gamma}\cos(\sqrt{\omega_0^2-\gamma^2}t+\theta_0)+c_1\cos(\Omega t+\theta_1) \tag{7}$$

where $c_0,\theta_0$ are arbitrary constants, and $c_1,\theta_1$ satifies

$$c_1 = \frac{1}{\sqrt{(\omega_0^2 - \Omega^2)^2 + 4\gamma^2\Omega^2}},$$
$$\theta_1 = crc\tan(\frac{-2\gamma\Omega}{\omega_0^2 - \Omega^2}) \quad (8)$$

Presuming $\omega = 30, \omega_0 = 10, \theta_0 = \pi, \gamma = 0.1, t = 5s,$
The result of simulating Eq (7) is shown in Fig.2 and the different damping coefficient vibration image of one-dimensional oscillator in Fig.3.

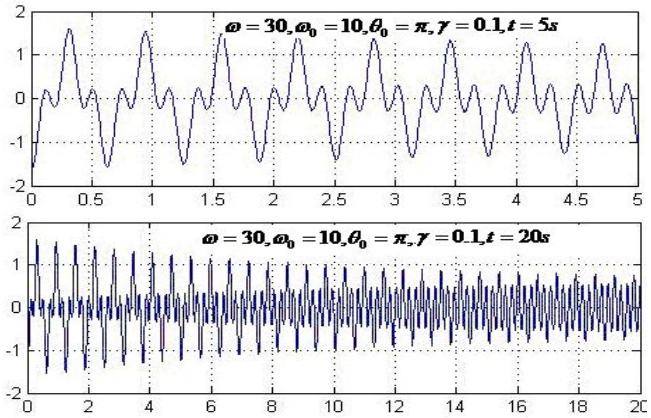


Figure.2 The damping vibration image of one-dimensional oscillator

Where the first containing attenuation index is the attenuation term, the second is the forced vibration term. When the time tends to be longer, the first item becomes zero and then the frequency of the harmonic oscillator is equal to the forced vibration frequency $\Omega$.

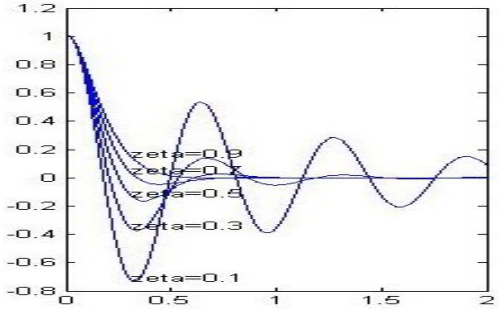


Figure.3 The different damping coefficient vibration image of one-dimensional oscillator

$$x(t) = c_1\cos(\Omega t + \theta_1) \quad (9)$$

If $\gamma, f(t)$ in Eq.(2) are zeros, that is, without dampingforce and external force, the equation becomes

$$\ddot{x}(t) + \omega_0^2 x(t) + \varepsilon\beta_0^2 x^3 = 0 \quad (10)$$

where the second is the nonlinear term. *Eq.*(10) can be used to describe many nonlinear vibration by the equation. To solve *Eq.*(10), we introduce the elliptic equation (where $A_j$ is constant)

$$y'' = A_0 + A_1 y + A_2 y^2 + A_3 y^3 \quad (11)$$

when $A_0 = A_2 = 0, A_1 = -\omega_0^2, A_3 = -\varepsilon\beta_0^2$, it has the same form with equation (10).

If $\varepsilon > 0$, the solution of the equation is Jacobi elliptic functions

$$x(t) = acn(\omega t, m) \quad (12)$$

$$\omega^2 = \omega_0^2 + \varepsilon\beta_0^2 a^2, m^2 = \frac{\varepsilon\beta_0^2 a^2}{2\omega^2}$$

If $\varepsilon < 0$, the solution of the equation becomes

$$x(t) = asn(\omega t, m) \quad (13)$$

$$\omega^2 = \omega_0^2 + \frac{1}{2}\varepsilon\beta_0^2 a^2, m^2 = \frac{-\varepsilon\beta_0^2 a^2}{2\omega^2}$$

This indicates no matter $\varepsilon > 0$ or $\varepsilon < 0$, the circular frequency of nonlinear vibration have relationship with amplitude. The circular frequencies of the above two cases could be calculated by the period of the Jacobi elliptic function, and results show that they are same, *i.e.*

$$\omega_e = \omega_0 + \frac{3}{8\omega_0}\varepsilon\beta_0^2 a^2 + O(\varepsilon^2 a^4) \quad (14)$$

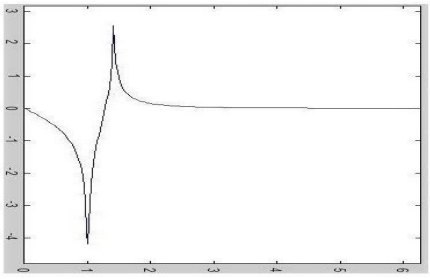


Figure.4 The nonlinear vibration image of one-dimensional oscillator

The conclusion shows that, the equivalent circular frequency equal to the sum of the natural frequency and the additional items caused by the nonlinear term. If $\varepsilon$ is minimal, the conclusion is in coincide with equation (4). In the cases of $m = 0.5$ ,the simulation was run fore equation (13),and the result is shown in figure 4. Nonlinear relationship of amplitude and $\omega$ is displayed obviously in the figure.

The equivalent equations of without damping force system described by equation (10) can be expressed as

$$\begin{cases} \dot{x}(t) = y \\ \dot{y} = -\omega_0^2 x - \varepsilon\beta_0^2 x^3 \end{cases} \quad (15)$$

And the phase trajectory on the phase plane (x,y) satisfies the equation

$$\frac{dy}{dx} = -\frac{\omega_0^2 x + \varepsilon\beta_0^2 x^3}{y} \tag{16}$$

The total energy of system stays a constant through the integral of *Eq*.(16)

$$H(x,y) = \frac{1}{2}y^2 + \frac{1}{2}\omega_0^2 x^2 + \frac{1}{4}\varepsilon\beta_0^2 x^4 \tag{17}$$

The system equation (15) can also be expressed in Hamilton canonical equation

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial y} \\ \dot{y} = -\dfrac{\partial H}{\partial x} \end{cases}$$

Although added the nonlinear term, the vibration system without forced and damping is still a conservative system.

When $\gamma, \varepsilon, f(t)$ are not zero, the equation (2) becomes nonlinear

$$\begin{aligned} &\ddot{x}(t) + 2\gamma\dot{x}(t) + \omega_0^2 x(t) + \varepsilon\beta_0^2 x^2 \\ &= A\cos\Omega t, (f(t) = A\cos\Omega t) \end{aligned} \tag{18}$$

the equivalent frequency can be introduced into *Eq*.(18) and it transformed to

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \omega_e^2 x(t) = A\cos\Omega t \tag{19}$$

Where $\omega_e = \omega_0 + \dfrac{3\beta_0^2 a^2}{8\omega_0}\varepsilon$ , the forced solution of equation (20) can be obtained when the time is long enough.

$$x(t) = a\cos(\Omega t + \varphi) \tag{20}$$

$$a = \frac{A}{\sqrt{(\omega_0^2 - \Omega^2)^2 + 4\gamma^2\Omega^2}},$$

$$\tan\phi = -\frac{2\gamma\Omega}{\omega_0^2 - \Omega^2}$$

When the forced vibration frequency and the inherent frequency is close, we define

$$\Omega = \omega_0 + \varepsilon\Delta$$

when $\Delta = 0$ , the system is resonant and the amplitude *a* would reach the maximum value. when $\Delta \neq 0$ , the magnitude of the amplitude would change nonlinearly along with the increase of $\Delta$ , which is nonlinear effect. If the forced vibration frequency varies great with the inherent frequency, the long-time solution of the forced nonlinear equation (2) according to (21) can be expressed as

$$x_0 = a\cos(\Omega t + \varphi)$$

$$a = \frac{A}{\sqrt{(\omega^2 - \Omega^2)^2 + 4\gamma^2\Omega^2}}, \tag{21}$$

$$\omega = \omega_0 + O(\varepsilon)$$

The solution can be regarded as a zero order approximate solution of equation (2). We define

$$x = x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots \tag{22}$$

where $x_1, x_2$ are the first approximation solution and the second approximate solution. Taking Eq.(23) into Eq. (2)

$$\ddot{x}_1(t) + 2\gamma\dot{x}_1(t) + \omega_0^2 x_1(t) = -\varepsilon\beta_0^2 x_0^3 \tag{23}$$

together with Eq. (22), (24), and the result is

$$\begin{aligned} &\ddot{x}_1(t) + 2\gamma\dot{x}_1(t) + \omega_0^2 x_1(t) \\ &= -\frac{1}{4}\varepsilon\beta_0^2 a_0^3[\cos 3(\Omega t + \phi) + 3\cos(\Omega t + \phi)] \end{aligned}$$

## III.CONCLUSION

Where there is a corresponding vibration due to the forced contains $\cos 3(\Omega t + \varphi)$ , that is to say, it will also produce the corresponding resonance phenomenon which is another important feature of nonlinear effects. The other resonance frequencies will be produced if considering the higher approximation.

## ACKNOWLEDGMENT

This work was supported by Zhoukou Normal University youth funding (No.zksyqn201329A)

## REFERENCES

[1] Dong-hu Deng,Zhang Qun,et al.The application of Duffing oscillators to micro-motion feature extraction of radar target under low SNR. Journal of Electronics & Information Technology, 2014(02):453-458.
[2] Ti-jun Li.The wave functions for a time-dependent harmonic oscillator Journal of Jilin Normal University (Natural Science Edition),2008(01):57-60.
[3] Fu-he Lu. Analysis and application for the nonlinear vibration of spring oscillator.Journal of Shanghai Second Polytechnic University, 2009(03):223-227.
[4] Jing-tao Wang， Min-xi Cheng， et al. Analyzing the damped vibration of spring oscillator on air track by software Tracker.College Physics, 2014(04):22-24.
[5] Rong-juan Zhou, Xiang-long Guo. Description of the x-t curve of the one dimension oscillator in the overdamping and critical damping states using microcomputer . Physics and Engineering, 2002(04):46-49.
[6] Guang-Tao Ding.On the first integrals of linear damped oscillators. Acta Phys. Sin. No. 6 (2013) 064501
[7] Shu-chuan Li. More discussions on oscillator motion.
[8] Journal of Sichuan Teachers College (Natural Science), 2002(03):307-309.
[9] Duan-zheng Yao. Method of mathematical physics. Beijing: Science Press,2010

# Applying Data Mining Technology to Solve the Problem of Traffic: A Case Study

Ke Zhu
Department of Educational Technology, Henan Normal University, Xinxiang, China
Email:zk@htu.cn

***Abstract*—Traffic rules assure the safety and efficiency of transportation systems. In this paper, we establish three models to analyze the performance of traffic rules.**
**Model deals with the traffic rules in single-lane. We simulate the vehicles movement and focus on analyzing the relationships of traffic flow, safety, and speed limits. In order to examine tradeoffs between traffic flow and safety, the role of speed limits under the rule of keep-right-except-to-pass, we propose a new model based on model 1. In addition to the driving rules referred in NS model, vehicles observe lane changing rules in a two-lane freeway. We then improve the model and apply it to countries where driving automobiles on the left is the norm.**

***Index Terms*—Traffic Flows, Statistical Analysis, Fuzzy comprehensive analysis**

## I. INTRODUCTION

Traffic rules assure the safety and efficiency of transportation systems. In countries where driving automobiles on the right is the rule, multi-lane free-ways often employ a rule that requires drivers to drive in the right-most lane unless they are passing another vehicle. Should new rules are applied, the traffic system can be better [1].

In order to examine tradeoffs between traffic flow and safety, the role of speed limits under the rule of keep-right-except-to-pass, we propose a new model (Two-lane overtaking model) based on model 1.In addition to the driving rules referred to NS model, vehicles observe lane changing rules in a two-lane freeway. We then improve the model and apply it to countries where driving automobiles on the left is the norm.

## II. METHODOLOGY

### A. Assumptions

- Assuming that the freeway we studied is straight and in good condition for moving on.
- The influence of bed weather on traffic is not taken into consideration.
- We assume that the vehicles moving on the freeway are in a good condition.
- The vehicles moving on the freeway don't distinguish each other from types.

Under the above and basic assumptions, we can set out to construct our model (show our approach in detail).

### B. The Single-lane NS Model

Cellular automata (CA) are discrete, dynamical systems that have proved useful both as general models of complexity and as more specific representations of non-linear dynamics in a variety of scientific fields. They are composed of a finite or denumerable set of homogeneous, simple units, the atoms or cells [2]. At each time unit, the cells instantiate one of a finite set of states. They evolve in parallel at discrete time steps, following state update functions or dynamical transition rules: the update of a cell state obtains by taking into account the states of cells in its local neigh *boyhood*. A large number of cellular constitute the evolution of a dynamic system by simple interactions.

The first application of the CA for simulation of traffic flows on streets and highways, which is called NS model, was introduced by Nagel K and Scheckenberg M [3].

NS model is discrete both in space and time. It regards a freeway as a one dimensional lattice grid on which vehicles moving in one direction.

The chain is divided into cells of length 7.5m.Each cell can either be empty or occupied by exactly one car.

Each vehicle is characterized by its current velocity $v$ which can take the values ranging from 0 to $v_{max}$. Here corresponds to a speed limit for all cars. $v(i; t)$ denotes the velocity of vehicle $i$ at time $t$. The maximum velocity is defined to be 5 cell length.

If $x(i; t)$ denotes the position of the vehicle $i$ at time $t$, the position of the car ahead of it at time t is $x(i + 1; t)$.The number of unoccupied cells between two consecutive vehicles is defined as gap.

The time step is taken to be 1$s$.

The following rules are applied in each iteration from $t \rightarrow t + 1$.It consists of 4 steps that have to apply at the same time to all the cars (parallel or synchronous dynamics).

*S1*: Acceleration the velocity of vehicle $i$ is increased by 1 cell-length.

$$v(i, t+1) = max\{v(i, t) + 1, v_{max}\}$$

*S2*: Deceleration the velocity of vehicle i is decreased by 1cell-length

$$v(i, t+1) = min\{v(i, t) - 1, gap\}$$

*S3*: Randomization the velocity of vehicle $i$ is decreased randomly by 1 cell-length with probability p.

$$v(i, t+1) = max\{v(i, t) - 1, 0\}$$

*S4*: Vehicle movement Each vehicle is moved forward according to its new velocity.

$$x(i,t+1) = x(i,t) + v(i,t+1)$$

We obtain the performance of single-lane NS model using cellular automata simulation program both in light and heavy traffic. The deceleration probability $p$ is set to be 0.3.

Taking tradeoffs between traffic flow and safety into consideration, we propose a safety coefficient. If the gap between two consecutive vehicles is greater than 7 cell-length, we define that the car behind is in a safe state. Otherwise the car behind is in a dangerous state. The safety coefficient corresponds to the ratio of vehicles in a safe state at a certain time.

As showed in Figure 1, the safety coefficient is changing with time variation. This is because the traffic flow is different at different time.

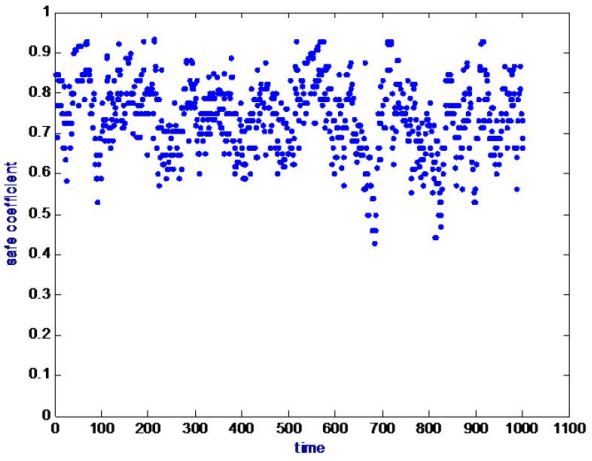



Figure 1. Safety coefficient at different time

By calculating and analyzing simulation data, we have the correlations of traffic flow and safety coefficient in Figure 1 . It indicates that the safety coefficient is changing with traffic flow variation. In fact, we find the simulation data fitting a function

$$y = -0.000017 \times x^5 + 0.001313 \times x^4 - 0.040152 \times x^3$$

therefore, we determine that under a certain level of traffic flow, the greater the traffic flow is, the smaller the safety coefficient is.

### C. Two-Lane Overtaking Model

We define a two-lane freeway as a two-dimensional lattice grid on which vehicles moving in one direction. And gap(i),gap(i, other),gap(i, back) respectively correspond to the number of unoccupied cells between the car i and the vehicle in front of it, the vehicle in front of it in adjacent lane and the vehicle behind it in adjacent lane.

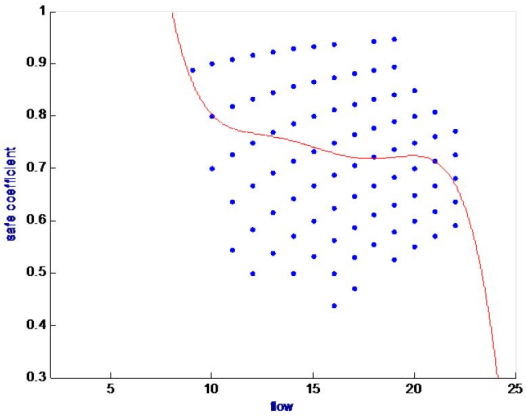



Figure 2. Correlation of traffic flow and safety coefficient

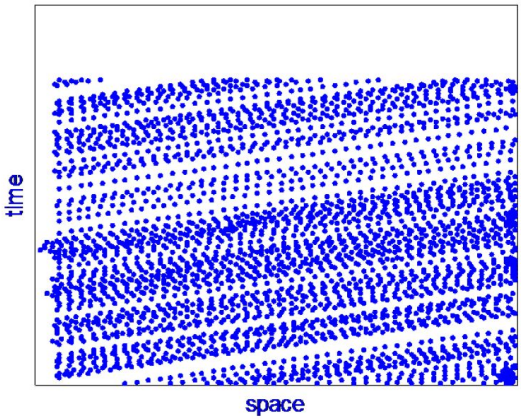



Figure3. Performance of space with time changing

In addition to the rules applied to NS model in each iteration, vehicles observe the following lane changing rules: Drivers drive in the right-most lane unless they are passing another vehicle, in which case they move one lane to the left, pass, and return to their former travel lane.

### D. Light Traffic Model

In order to examine tradeoffs between traffic flow and safety and the role of under- or over-posted speed limits, we simulate the vehicles movement process in a two-lane freeway. The value of 0.7 is given to lane changing probability $q$ by searching literature.

Traffic density can reflect the degree of freeway safety to a certain extent. Safety factor generally decreases with traffic density increasing. Hence we measure the traffic safety degree by traffic density.

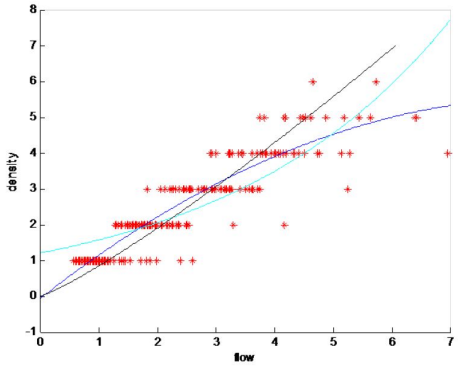



Figure 4. Correlation of traffic flow and traffic density

It can be easily seen that traffic flow and traffic density appear a positive relationship in Figure 4.When traffic

flow increases, the traffic density also increases. We obtain three fitting curves by logarithmic fitting, quadratic fitting and exponential fitting:

$$f1: y = e^(0.1515 + 0.8490 \times log(x))$$

$$f2y = 0.3003 \times x^2 - 5.2113 \times x + 76.0365$$

$$f3y = 74.9239 \times e^{0.0550}$$

Since we believe that the safety factor generally decreases with traffic density increasing, the safety factor decreases with traffic flow increasing.

Speed limit also plays an important role in traffic simulation. The Figure 5 indicates traffic density generally decreases with velocity limit increasing so we can consider the safety factor decreases with velocity limit increasing. We obtain three fitting curves by logarithmic fitting, quadratic fitting and exponential fitting:

$$f1: y = e^(0.1515 + 0.8490 \times log(x))$$

$$f2y = 0.3003 \times x^2 - 5.2113 \times x + 76.0365$$

$$f3y = 74.9239 \times e^{0.0550}$$

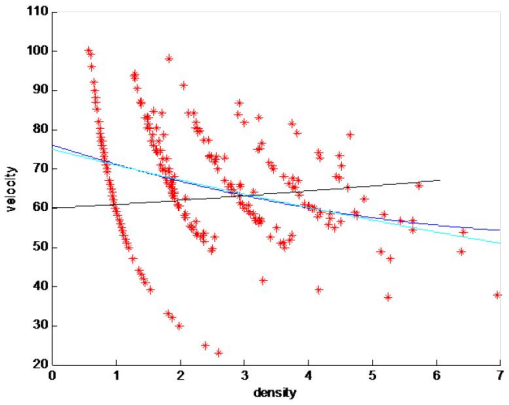



Figure 5. Correlation of traffic density and velocity

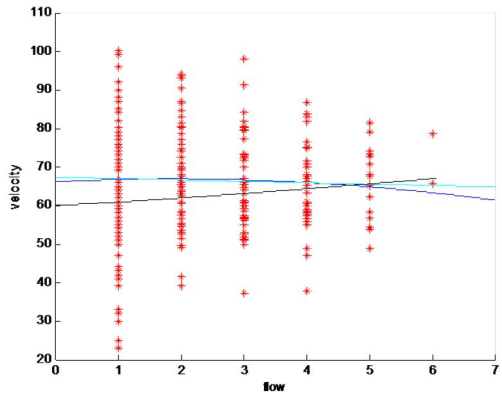



Figure 6. Correlation of traffic density and velocity

The Figure 6 apparently shows that the relationship of speed limits and traffic flow is quite small. This suggest that the traffic flow is not sensitive to speed limit in light traffic. We obtain three fitting curves by logarithmic fitting, quadratic fitting and exponential fitting:

$$f1: y = e^(0.1515 + 0.8490 * log(x))$$

$$f2: y = -0.2092 \times x^2 + 0.7833 \times x + 66.1212$$

$$f3: y = 67.3324 \times e(0.0550)$$

### *E. Heavy Traffic Model*

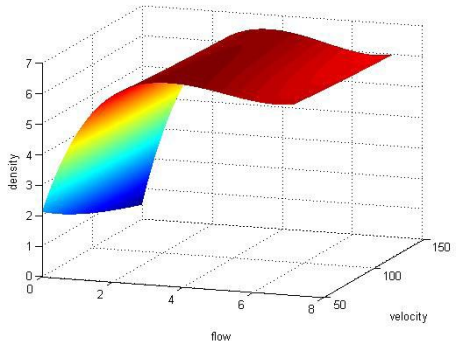



Figure 7. Relationships among traffic flow, traffic density and velocity

By simulating the process in a two-lane freeway in heavy traffic, we obtain the relationships among traffic flow, traffic density and velocity.

When traffic flow increases, the traffic density also increases by quadratic curve. And when the traffic flow increases to a certain extent, the density gradually achieve smooth.

In general, the relationship between velocity and traffic flow is quite small in heavy traffic.

The traffic density increases with velocity increasing. But the increasing trend is gradually slowing down.

Comparing the NS single-lane model with two-lane overtaking model, we consider that the keep-right-except-to-pass rule has a better performance on enhancing traffic flow to a certain extent. We determine the ordinate velocity in figure 7 reflects the freeways capacity. When the traffic flow on the freeway increases, both the single-lanes capacity curve and the two-lanes capacity curve decrease. But single-lanes capacity is more sensitive to traffic flow increasing. So we believe the keep-right-except-to-pass rule has a better performance on enhancing traffic flow comparing to the single-line rule.

Using the Computer Simulation, we obtain the data of traffic flow, traffic density and velocity under the rule of driving automobiles on the left. Then we trace the points under two different rules. By the method of fitting we get the relationship between any two parameters.

As described in the above tables, the same fitting method gets similar curves. It indicates that the model we establish also applies to the countries approximately where driving automobiles on the left is the norm.

## III. Conclusions and Further Explanations

Expected driving speed: in the free driving state, vehicles suffer little constraints. In this condition, drivers can get the speed they want. They change the driving speed to achieve a certain purpose, such as shortening the travel time, enjoying the driving process and so on. Expected driving speed is affected by the driver's desire, the fatigue degree, driving purpose and so on. Therefore,

the desired speed varies with different drivers. In our above models, we assume speed obeys normal distribution. When the speed is higher than the desired speed, vehicles slow down, and when the speed is lower than the desired speed, vehicles accelerate. However, in intelligent control system, we consider that the expected driving speed is the highest limited speed.

The speed limit: studies show that under The Keep-Right-Except-To-Pass Rule, the relationship between accident rate and speed is a U-shaped curve. When the speed is close to the average speed, the accident rate is lowest. The greater the difference between the real speed and the average speed is, the bigger the accident rate is. In our models, in order to improve safety, we set the under- or over-posted speed close to the average speed. In addition, the dispersion degree between the real speed and the average speed is very small. However, in the intelligent control system, the relation between accident rate and speed is weak. Therefore, in this system, the under- or over-posted speed limits on the highway have little influence on the safety.

ACKNOWLEDGMENT

This work is supported by soft-scientific project of Henan Province (No. 142400410425)

REFERENCES

[1] Nagel KScheckenberg M.ACellular Automaton model for freeway trafficJ.JPhys1992I-22221- 2229
[2] Daganzo C F.A behavioral theory of multi-lane traffic flow.Part 1:Long homogeneous freeway section. Transportation Research B.2002
[3] Jean-Louis Martin, Relationship between crash rate and hourly traffic flow on interurban motorways. Accident analysis and prevention. 34(2002)

**Zhu ke:** born in 1982, His research interests include learning analysis, Computer education.

# Time Sensitive Proximity Based Routing Protocol in Opportunistic Networks

Li Liu
Shandong Jiaotong University, Jinan 250023, China.
liuli_free@126.com

***Abstract*—Socially-aware Opportunistic Networks are used in intermittent connection network consisted of mobile devices by use of short range transmission technology. In this paper, we present a time sensitive proximity based routing protocol (TSP) and utilize the irregularity of mobility about physical proximity and time to predict the future meet opportunity. We establish experiment and the simulation results shown that the efficiency of TSP obtains higher performance.**

***Index Terms*— Opportunistic Networks, Physical Proximity, Routing Protocol.**

## I. Introduction

With the development of wireless transmission technology, more and more mobile devices can exchange messages through short range transmission technology such as Bluetooth and WiFi. On the same time, more and more mobile devices are used and carried by people. However, in this scenario, an end-to-end path is difficult to maintain due to the mobility of devices. In these intermittent connect environment, socially-aware Opportunistic Networks [1] then emerge, which utilize opportunistic contact opportunity to deliver messages using store-carry –forward fashion. When a source device wants to deliver a message to a destination device, several intermediate devices may be selected to help the forwarding process hop by hop.

Therefore, the prediction of meet probability is the key problem to obtain higher delivery performance. Recently, the human's social relationships attract most of researchers' attention. Many recent researches are inspired by the social properties such as [2-4].

Physical proximity is an important factor for implementing the direct delivery. Since only when two devices are closely enough, the transmission can be happened. Furthermore, the human's mobility is regular, especially in work days. Thess properties can be helped to predict the meet probability between nodes.

In this paper, we present a time sensitive proximity based routing protocol (TSP), which is based on the regular proximity information of human's mobility to predict the future meet opportunities. We establish simulation and evaluate the performance of TSP through comparing to Epidemic routing protocol and PROPHET routing protocol. The simulation results demonstrate the higher efficiency of TSP.

The rest of this paper is organized as follows. Section 2 provides a review of related works on socially-aware Opportunistic Networks. Section 3 describes the system model and implementation of TSP Routing. We simulate and evaluate the performance of TSP Routing using real trace data in Section 4. Finally, conclusion is presented in Section 5.

## II. System Model and Implement

### *A. System Model.*

In this section, we give a detail description for the system model. We assume there are N mobile devices and they transmit with each other through Bluetooth. And the mobility of mobile devices is regular through time. We can imagine a campus scenario. Campus has scheduled time table. Students may have breakfast, lunch and dinner at similar time and fixed location during the workdays.

In campus scenario, mobile devices record the proximity information at different time period. We divide the day time into 12 parts from 7:00am to 7:00pm in one hour. The night time is discarded because the mobility is rare in night. When two devices meets, they record the contact informaiton with each other. The contact information includes nodes' ID and nodes contact times for this time period.

### *B. Implement*

In TSP, mobile devices record and maintain proximity information for all contact nodes. When two nodes (called A and B for simply description) contact each other. They update their proximity information firstly. Then they exchange the message list to choose the better forwardre. For the message in one mobile device, TSP will compute the meet probability of the destination by comparing the similarity for proximity information. The TSP routing protocol are presented as follows, which consists of 4 steps.

1) The mobile devices record and maintain the proximity information as they are moving.

2) When two mobile devices are in their transmission range. They exchange their physical location information firstly.

3) Then two devices exchange the message list with each other. Firstly, node A checks each message's destination and computes the meet probability (PA) between A and the destination. Then A compares PA to PB in message list. If PA > PB, the message will be delivered by A. Otherwise, the message is stayed in B. B

has the similar process with A. The meet probability is caculated through comparing the similarity degree of proximity information between the node to message's destination. If the node has higher similarity degree with destination node, they will have higher probability to meet in future.

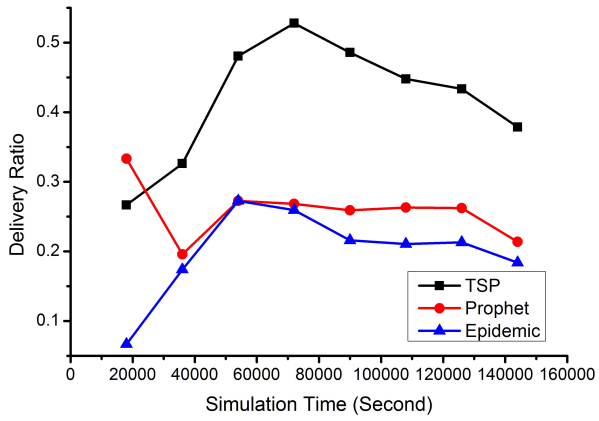
4) Start the transmission process until beyond the transmission range.

The value of meet probability between A and destination is the sum of number of same location at corresponding time.

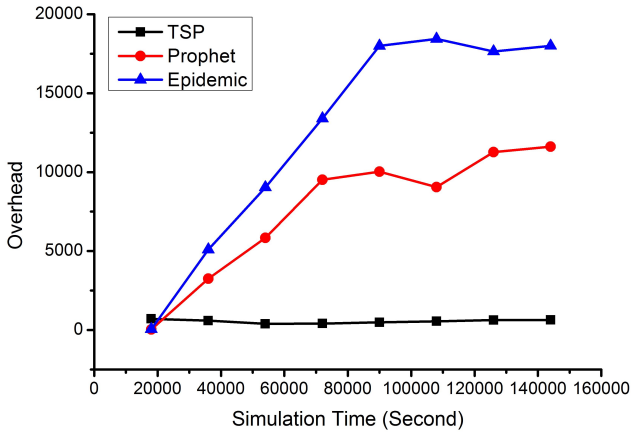
## III. SIMULATION

In the simulation, we use real data set SIGCOMM2009 [9] through ONE simulator [10]. The messages with 8 hours TTL are generated randomly.

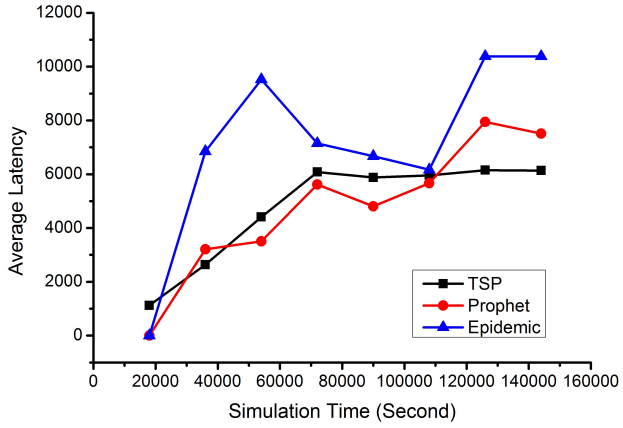
In the simulations, three performance metrics are evaluated from four aspects: delivery ratio, average latency, overhead ratio and average hop count.



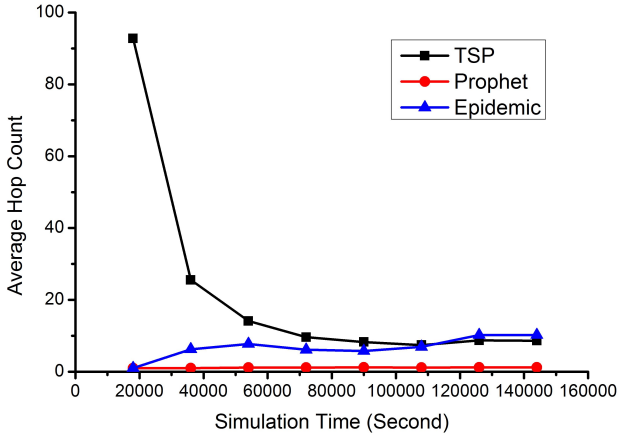

(a) Delivery Ratio




(b)Overhead




(c) Average Latency




(d) Average Hop Count

Fig. 1 Efficiency Comparison

We compare the efficiency of TSP with two famous routing protocols: Epidemic and PROPHET. The simulation result is illustrated in Fig 1. As we can see, Fig 1 shows comparisons from delivery ratio, overhead ratio, average latency and average hop count respectively. As Fig 1 shown, the performance of TSP outperforms Epidemic and PROPHET highly. For example, in 18 hours, TSP forwards 53.45% messages with overhead ratio of 78.2, average latency of 6122 and average hop count of 8. While the delivery ratios of Epidemic and PROPHET are 24.6% and 26.71% respectively with overhead ratio of 1346 and 824 respectively. The average latency of Epidemic and PROPHET are 7320 and 5983. The average hop count of Epidemic and PROPHET are 6 and 1.

## IV. CONCLUSION

In this paper, we present a time sensitive proximity based routing protocol (TSP) in socially-aware Opportunistic Networks. In TSP, each mobile device records and maintains proximity information according fixed time period. And based this information, TSP predicts the meet probability and selects better forwarder in order to improve the efficiency. The simulation refers

that TSP obtains higher performance comparing with Epidemic and PROPHET.

## ACKNOWLEDGEMENTS

This work was financially supported by the Shandong Jiaotong University Science Research Foundation (Z201305).

## REFERENCES

[1] M. Conti, S. Giordano, M. May, and A. Passarella, "From opportunistic networks to pportunistic computing," Communications Magazine, vol. 48, no. 9, pp. 126–139, 2010.

[2] H. Pan, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," Mobile Computing, vol. 10, no. 11, pp.1576–1589, 2011.

[3] F. Li and J. Wu, "Mops: Providing content-based service in disruptiontolerant networks," Distributed Computing Systems, 2009. 29th IEEE International Conference on, pp. 526–533, June 2009.

[4] L. Feng and W. Jie, "Localcom: A community-based epidemic forwarding scheme in disruption-tolerant networks," 6th Annual IEEE Communications Society Conference on, pp. 1–9, 2009.

[5] A.-K. Pietil¨ainen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "Mobiclique: middleware for mobile social networking," Proceedings of the 2nd ACM workshop on Online social networks, pp. 49–54, 2009.

[6] A. Ker¨anen, T. K¨arkk¨ainen, and J. Ott, "Simulating mobility and dtns with the one," Journal of Communications, vol. 5, no. 2, pp. 92–105, 2010.

# Study on E-commerce Transactions Credit Rating of China's Foreign Trade Enterprise

LI Hanwei; LIU Meng
Shanghai University Of Engineering Science, School of Management, Shanghai, China

***Abstract*—This paper first expounds the importance of China's foreign trade enterprise e-commerce transactions credit rating, then presents the principle, method and procedure of e-commerce transactions credit rating in China's foreign trade enterprise, the article finally has carried on objective evaluation to the rating method, and gives some reasonable Suggestions. Research ideas and methods of this article has certain characteristics and innovative, has strong reference and application value.**

***Index Terms*—Foreign Trade Enterprise , E- commerce Credit Rating , Credit Rating**

## Ⅰ. INTRODUCTION

As China's foreign trade enterprises face the external environment deteriorating, there are a large number of foreign trade enterprise closed down in Shanghai. The reasons are not only the influence of the global economic recession, but also the competitive ability of foreign trade enterprise is not strong. It is worth noting that under the radiation of global financial crisis, according to the data of the Shanghai government research center, those small and medium-sized foreign trade enterprises which use B2B under the world financial crisis are much better in development situation than those who are not using e-commerce, the survival rate is as high as 4.4 times or more. Based on that background, because of the low cost and high efficiency to carry on products promotion , using e-commerce trading platform at that time became the first selection of small and medium-sized enterprises. This is because the foreign trade enterprises can reduce cost, effective use of space and time through the China has the foreign trade enterprise credit rating standards and measures for its implementation, but under the background of the electronic commerce as a way of foreign trade enterprises transactions. To evaluate the credit rating on foreign trade enterprises, there is no related research on foreign trade enterprise electronic commerce trade credit ratings.

Therefore, in order to project the overall level of the foreign trade enterprise electronic commerce, to improve the competitive ability of the foreign trade enterprises ,it is urgently need to rating the e-commerce credit of foreign trade enterprises, at the same time ,in the mastery of the foreign trade enterprise overall credit information, it also can be used to urge on some enterprises whose credit ratings are not high, to improve the overall credit level of foreign trade enterprises, then improve the competitiveness of the foreign trade enterprise in Shanghai.

## II RESEARCH GOALS

Aiming at the shortages of the current credit rating model under the electronic commerce transaction mode, introducing the network information data mining methods and introducing the non-financial factors into the evaluation model, considering comprehensive the qualitative and quantitative, financial and non-financial factors, putting forward more accurate Multi-level fuzzy evaluation model for small and medium-sized foreign trade enterprise credit rating, and using the foreign trade enterprise data of world, rating foreign trade enterprise credit under the electronic commerce transaction mode for China's foreign trade enterprises, to provide decision-making information for government management departments, being an example with good credit records for foreign trade enterprises ,encouraging foreign trade enterprises with better credit continue to maintain the good credit, to urge on those foreign trade enterprises with poor credit to improve the foreign-trade credit, to improve the overall competitiveness of China's foreign trade enterprises.

## III. THE PRINCIPLES OF E-COMMERCE TRANSACTIONS CREDIT RATING OF FOREIGN TRADE ENTERPRISES

Establishing e-commerce credit rating index system , the indicators must be organic coordination, forming a system, neither repeat with each other, nor contradiction; At the same time, The calculation of indicators and rating methods must be scientific, should have certain basis, to fully draw lessons from foreign experience about the development of the enterprise and individual credit rating. The content of the e-commerce credit rating index system should fully reflect all the elements that affect credit conditions of the rating object, not only consider the enterprise performance in the past, but also predict the development tendency in the future; Not only consider the hardware condition of the enterprise, but also study internal credit management condition. Only in this way can we achieve the comprehensive rating requirement .We can not make the conclusion of credit rating only depend on minority rating indicators, so that it is easy to produce the rating inaccurate error. E-commerce credit rating index system must be targeted, The index system should be different between different rating object and rating purpose. Traditional credit rating

in China at present are classified into securities rating, loan enterprise rating and specific credit relations rating . But for e-commerce, we should take into account the characteristics of different market subjects, such as B2B, B2C, individual consumers, platform providers, financial institutions, etc. E-commerce credit rating must abide by the relevant state policies, laws and regulations, The index system should reflect the national macro policy guidance, Some economic benefit indicators and risk supervision indicators that have standard values of state provisions, must meet the requirement of regulations. The establishment of the e-commerce credit rating index system, must conform to the objective facts, and can correctly reflect the real appearance of the rated objects credit rating, index system and calculation method cannot be partial to any rating object. The attitude of rating agencies and staff must be fair, the rating must be objective and based on facts. The establishment of the e-commerce credit rating index system must be practical and easy to operating and program calculating. It should confirm to national condition of China, and possess Chinese characteristics, refer to international conventions, considering the connection with the world in future, so that facilitating the integration of international credit rating.

## IV THE RESEARCH METHODS AND STEPS

### *A. Analysis of the shortcomings and disadvantages of the trademark evaluation method and the trademark WALE measurement of the existing foreign trade enterprise credit rating*

Trademark evaluation methods are too simple, this method only depends on some isolated factors to add and subtract, without considering different weights between various factors, and most of the evaluation factors adopted by trademark evaluation methods through the perspective of the external environment of enterprise management, there's no involved enough develop factors of enterprise itself, It is not enough for the enterprise credit rating, at the same time ,it is also not easy to assess the enterprise credit.

Although WALE credit evaluation model is suitable for most of the foreign trade enterprise credit assessment, but does not take into account the particularity of foreign trade enterprises under e-commerce transaction mode ,without targeted enterprise credit rating under the background of e-commerce transactions , in order to avoid this defect and to improve the accuracy of the credit rating, this paper intends to using the more active current research theory---- fuzzy comprehensive evaluation model in the aspect of credit rating model, And introduced the factors affecting China's B2B e-commerce credit , the credit rating system can according to the development of market economy and the change of the objective environment to adjust their function, and to continuously improve the accuracy of credit rating.

### *B. To determine the enterprise credit rating factor set*

Through the questionnaire survey method, conducting statistical analysis and factorial analysis ,then combining with the method of data mining, comprehensive considering the financial factors and non-financial factors, qualitative factors and quantitative factors and identify factors and uncertainties, integrating the concept of fuzzy mathematics, to determine the foreign trade enterprise credit rating factor set. For the financial factors, considering the financial factors of foreign trade enterprises, such as the amount of customs declaration , the tax amount, scale and monthly profit margins, Non-financial factors mainly are the monthly number of import and export, information level and the customs rating; Quantitative factors can accurately measure data; Qualitative factors mainly consider public opinion and company between evaluation and network comments, etc.; Deterministic factors mainly collected primary data, or the processed data; Uncertain factors are mainly the change of the macroeconomic environment, as well as emergency impact on their own performance of foreign trade enterprise.

### *C. Determine the comment set*

Comment set is a collection of ranked elements divided by the results of assessment, no matter how many hierarchy of index system, there is only one comment set. Comprehensive study of various current credit rating classification method, this paper argues that five division method is more suitable for the credit rating of China's foreign trade enterprises (mainly small and medium-sized enterprises), therefore, in this paper, small and medium-sized enterprise's credit rating can be divided into five grades. Respectively be represented by letters AAAA, AAAA, AAA, AA, A, the meaning are outstanding, good, general, poor and very poor. Thus determine the evaluation set Z = (AAAAA, AAAA, AAA, AA, A).

### *D. Establish weight vector*

According to the method of decision analysis, to determine the weight vector of index factors. Taking the expert scoring method to determine the index weight vector , and integrating the AHP analysis method, then can be relatively easy to determine the weight of each index, the determination of these weights combined the advantages of qualitative and quantitative, possessing strong practical availability.

### *E. Determine the membership degree of Each indicator in index layer*

According to the ideas of the fuzzy clustering, to calculating memberships of qualitative index and quantitative index. In light of that there are so many qualitative index in foreign trade enterprise credit rating system, such as rating between enterprises, network consensus and the information level ,if conduct clear attribution judgment, Rating work seem not enough meticulous, even very absolute and arbitrary, it is difficult to be accepted by the general public and related foreign trade enterprise, therefore, adopting the method

of fuzzy clustering to calculating membership degree is more practical.

*F. Multistage fuzzy comprehensive evaluation*

It includes the primary fuzzy comprehensive evaluation, the secondary fuzzy comprehensive evaluation and the advanced fuzzy comprehensive evaluation. This classified, hierarchical fuzzy comprehensive evaluation is to consider China's foreign trade enterprise in different stage, it possesses strong complexity, Plus of the macro environment of China's rating tolerance to foreign trade enterprises, in different stages of foreign trade enterprise rating popularity ,respectively take the primary fuzzy comprehensive evaluation, the secondary fuzzy comprehensive evaluation and advanced fuzzy comprehensive evaluation ,to make foreign trade assessment system seem more flexible and more convenient to promoting.

*G . Synthesize three-level calculation results, use the data of foreign trade enterprises, make the comprehensive credit rating of foreign trade enterprises in China.*

The source of the data can be divided into four aspects, one is the government channels, mainly including the customs, statistics bureau, industrial and commercial bureau, department of foreign trade and economic cooperation; The second is to provide a port that allows businesses to report their own data, if enterprise fill in some false data, they need to be punished until be kick out the rating ranking, the popularity of the rating system constantly improve, enterprises themselves to fill in the data will be more and more high; The third is the Internet channel, it need to collect and text analysis for Internet information, and to filtering and data mining, then to extract useful information; The fourth is investigation channels, through enterprises interview and questionnaire survey, to obtain some first-hand on quantitative data and practical and useful information.

## V THE EVALUATION OF RESEARCH METHODS

The research methods of this article considers the influence factors of e-commerce transaction mode to foreign trade; comprehensive Considering the financial indicators and non-financial indicators; Considering the qualitative indexes and quantitative indexes; Considering the certainty indexes and uncertainty indexes. Research method is based on the mature research methods in the foreign trade enterprise credit evaluation application fields of applied research breakthrough, with technical feasibility; China's customs department has the rich data of foreign trade enterprises, Overall research method has strong practical maneuverability, and has certain theory value and practical value.

## VI SUGGESTIONS FOR CHINA'S FOREIGN TRADE ENTERPRISE ELECTRONIC COMMERCE CREDIT RATING

(1) First of all, It is imperative to ratings for China's foreign trade enterprise e-commerce transactions, the government should actively participate in. Because, with improving China's international status and increasing international level, in order to enhance the competitiveness of the whole of China's foreign trade enterprises, we need to monitor and rating the e-commerce transactions, As China needs fund rating companies to rating the fund company .,It is indispensable. The government especially the customs should vigorously promote this work, and to guide and help, to provide data support, make it developing faster.

(2) As the rating foreign trade companies, should be gradual to develop their business. First of all, rating foreign trade enterprises of some familiar industry in an area, when the fame and recognition become higher, And then gradually expand from the area to the country, to expand from a industry into all industries , so it can carry out more smoothly.

(3) Selection of rating index should be reasonable, timely, and advancing with the times. when start data rating, we should choose some data available and reasonable indicators. With the development of the business and the need of reality, the index selection continue to increase and improve, in order to keep pace with the times, to design more accepted and comprehensive foreign trade enterprise credit rating system.

## REFERENCES

[1] Du Peng. Trying to explore a scientific evaluation system [N]. China securities journal, 2000-8-10.
[2] Zhiyong Li, Shun-hong Wang. Based on the fuzzy comprehensive evaluation method for supplier selection. Journal of Beijing university of technology, 2004, (3)
[3] Lei Bing. Design and development of electronic commerce credit evaluation system [J]. Microcomputer information (control integration), 2007, 23 (3) : 1732175.
[4] Qing-guo Ma .The reality of development of Chinese e-commerce paths: analysis and empirical [J]. Management world, 2004(02).
[5] Piao C H , An J , Fang M Q. Research on credit evaluation model and algorithm for C2C E2commercewebsite[J] . Journal of Information , 2007 , (8) : 1052107.
[6] Tian B , Tan Z. Study on the t rust model of ecommerce[J ] . Journal of Information , 2007 , (4) : 42-44.
[7] Wang Ting , Qu PengZhi. B2B e-commerce mode to promote our country the establishment of the enterprise credit management system [J]. Journal of business research, 2005 (21).
[8] Wei Gong. Famous overseas fund rating system introduction [J]. China financial city, 2004, (1)
[9] Yang Yan. The idea of China's e-commerce and credit information management system [J]. Journal of intelligence science, 2005(02).
[10] Yu Shouxi, Li-yan Han. The research of securities investment fund performance evaluation system [J]. Journal of capital university of economics and trade, 2004,
[11] Lin-feng Zhao, Kai Li. To build China's multi-level fund evaluation system [J]. Journal of northeastern university (social science edition), 2004, (5).

# Software Applications in Data Acquisition and Analysis of Battery Parameters Measurement

ZHANG Li-guo; Gao Jing
Northeast Petroleum University Qinhuangdao Branch , Qinhuangdao ,china
Nai_fan@163.com

***Abstract*— This paper builds a battery charge and discharge experiment platform, using the data acquisition board to complete the acquisition, experimental control and data communication. PC software and SCM build bridges through serial communication. The computer puts a large amount of data information through software to complete the storage and analysis. This paper describes the use of VB serial communication method, using Labview 2D/3D data visualization capabilities for data analysis and visualization software Matlab, intuitive analysis of the collected data. Achieve a number of software in the process of data collection and analysis of integrated use. This paper also describes the feasibility and practicality of experimental data analysis to verify meter program.**

***Index Terms*—Data collection, VB, data visualization analysis, Labview**

## I. INTRODUCTION

Computer powerful data processing capability is recognized. According to statistics, more than 80% of the computer is mainly used for data processing. Under the research field, requiring a large number of experimental records, the manual is a huge workload can think of, this time, the application of computer data processing capability is particularly important.

This paper describes the computer software VB, Labview and Matlab applications in the data processing and analysis. VB [1] has a rich database management capabilities, Labview [2] with data analysis and visualization tools to control functions, Matlab [3] has a numerical computing and 2D/3D data visualization capabilities. These software features can play to their strengths role in the data collection process, the use of serial communication technology applied to the battery parameter measurement data acquisition, integrated application software, not only to achieve the experiment saving time, but also to increase the data collection the amount. The use of software played a crucial role for study analyzes.

## II. BATTERY PARAMETERS COLLECTED EXPERIMENTAL DESIGN

### *A. Battery parameter detection realization diagram*

Figure 1 is a schematic diagram of the experimental battery charge and discharge. Battery parameters can be real-time monitoring and control. Charging Experiment power is added, discharging test power is removed. PLC is used to control the overall time of the experiment. Charge-discharge process was monitored to collect battery abnormality. Acquisition board control relays disconnect the charging and discharging circuit. Acquisition board functions can realize data acquisition and data transmission.

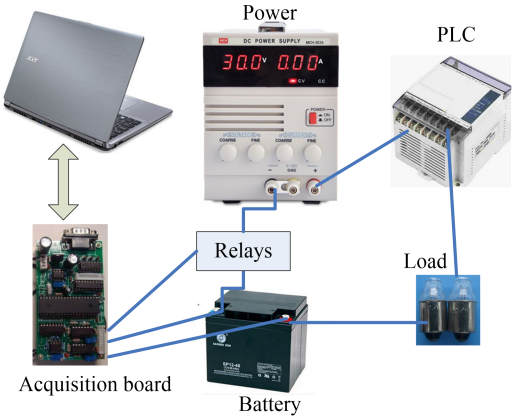



Figure 1 Block diagram of switching power supply

### *B. Experiment Principle*

This experiment is designed to measure the main achievement of the battery terminal voltage. Charge-discharge process of the battery terminal voltage measurement and the recording line, and then analyzes the power storage capacity. Data from the acquisition boards temporarily stored in the microcontroller, then transferred to a PC for storage and analysis via serial communication.

## III. SERIAL COMMUNICATION AND ITS IMPLEMENTATION IN THE SOFTWARE PLATFORM

### *A. VB serial interface. Labview serial port*

Serial communication program design to achieve is to use MSCOMM control that provides a lot of convenience operation properties and methods. They can easily use to achieve your goal first initialize the serial port, such as port number, baud rate and other properties, and then open the ports, uplink data read by the receive buffer, line by sending a buffer to write data. Finally, to reflect the arrival and sending data process through event-driven, in addition to errors occurred in the communication process can also be managed by COMMEVENT property.

Using MSCOMM control is mainly through the event to handle the interaction of serial port. That is, ONCOMM event of control will capture or handle these communications events when data arrives. The ONCOMM event can also be used to capture and handle communication errors. In practical applications, an MSCOMM control corresponds to a serial port, so if you want to deal with more than one serial port, then there must be a corresponding number of the corresponding controls. It can easily receive and transmit serial data buffer through the control of input and output attributes MSCOMM. The following code is VB to realize the serial communication program.

```
Private Sub Form_Load()
    Dim ReadsStr As String
' Save input substring buffer
        MSCOmm1.CommPort=1 ' Select the serial
port    MSCOmm1.Settings="9600,N,8,1"
' Communication settings
        MSCOmm1.Portopen=True  ' Open the serial
port
        MSCOmm1.Output="AT"+Chr$(13)
' Send command
    Do                      ' Waiting for data to return
to port
        DoEvents
    Loop Until MSCOmm1.InBufferCount>=2
    ReadStr= MSCOmm1.Input
' Response from the serial port
    ……
    MSCOmm1.PortOpen=False ' Close the serial port
End Sub
```

*B.Labview serial interface*

Labview is used to realize visual analysis of data acquisition. Labview software communicates with the serial Instruments via the VISA. VISA is a standard I/O application programming interface, mainly used in instrument programming. It provides the user with a separate set of standards I/O underlying function and it can easily call. In this experiment, the voltage signal to be transmitted to a number of decimal places. It is in hexadecimal form of data exists in the memory. Serial data is transmitted in the form of data packets for transmission, according to the serial communication protocol packets received first byte separation, and then were converted to a string of characters display, the resulting value is the actual physical quantities. We can combine the above features of the results and VISA serial functions. Program code is designed shown in the flowchart in Figure 2.

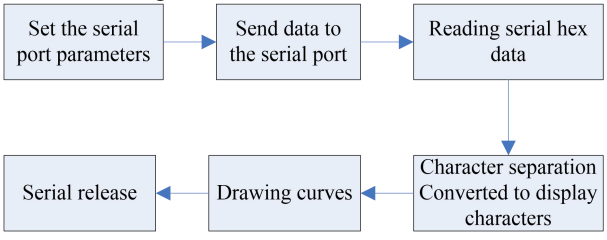



Figure 2 Serial communication program flow chart

Figure 3 is a front panel serial communication program. Dynamic parameter curve shows a chart of the Chart control, the waveform display as an analog oscilloscope form. Voltage data sampling interval according to the dynamic real-time curve display on the chart to the left of the chart area to send data command and data reception area. Panel also designed the serial port selection function, the baud rate setting function, clearing the data, erasing the curve and off serial port etc.

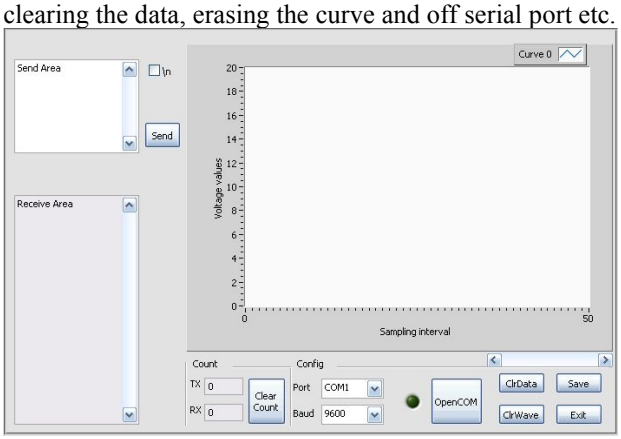



Figure 3 serial communication program panel

## IV. ANALYSIS OF EXPERIMENTAL DATA

*A. Discharge voltage log analysis*

Experimental measurement object use the battery for the San Yang SSP12-7 (12V/7.0Ah). Experimental conditions, the battery with the charging and discharging current 0.55A discharge cutoff voltage is 10.8V, Figure 4 is a 2D drawing using Matlab, some experimental data in Fig discharge every 30 minutes measure the battery terminal voltage.

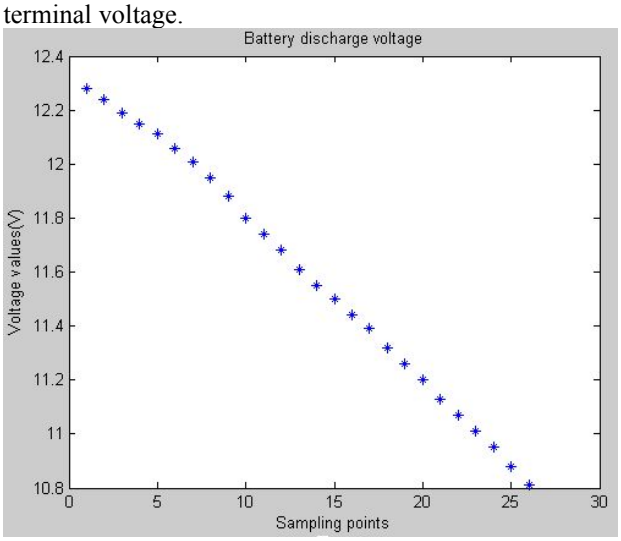



Figure 4 Discharge voltage analysis chart

As can be seen from Figure 4, As can be seen from Figure 4, the uniform decrease in the battery voltage during the constant current discharge. The discharge end of the actual discharge capacity of the battery is about 7.43Ah. For its full capacity of the battery discharge process is to start slow discharge voltage drop, the discharge end of the voltage drops rapidly. Drawing can be more intuitive use Matlab to analyze data

*B. Record discharge capacity analysis*

The use of batteries is the San Yang SSP12-7 (12V7.0Ah). Experimental conditions, the battery with the charging and discharging current 1.00A discharge cutoff voltage of 10.8V. Table 1 shows a battery multiple discharge experiments.

TABLE 1 BATTERY CAPACITY MEASUREMENT

| No. | Measured current (A) | Discharge time (h) | Battery capacity (Ah) |
|---|---|---|---|
| 1 | 1.00 | 6.75 | 6.75 |
| 2 | 1.02 | 6.65 | 6.78 |
| 3 | 1.03 | 6.57 | 6.77 |
| 4 | 1.08 | 6.20 | 6.70 |
| 5 | 1.10 | 6.00 | 6.60 |
| 6 | 1.05 | 6.33 | 6.65 |
| 7 | 1.03 | 6.60 | 6.80 |
| 8 | 1.09 | 6.12 | 6.67 |
| 9 | 1.08 | 6.24 | 6.74 |
| 10 | 1.05 | 6.32 | 6.64 |

Each discharge time of at least 14 hours, this process requires a large amount of battery voltage data recording, while the battery capacity is obtained by calculation. By using VB software fully solved the problem. Measurement data measurement circuit by means of SCM help, real-time data is sent using the serial port of the PC, PC machine using VB powerful database software for data storage and processing. Thus by VB software interface design, can easily be found in the required data.

As shown in Figure 5 using Matlab software to draw 3D view, more intuitive Table 1 shows the change in the three groups of data.

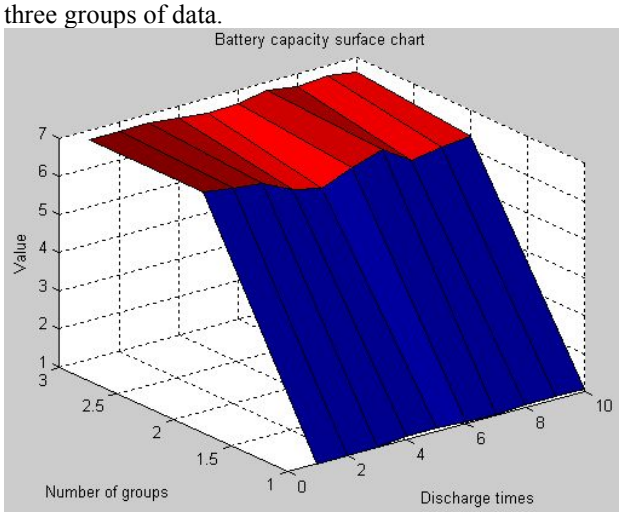



Figure 5 Record discharge capacity analysis chart

## V. CONCLUSION

Establish the experimental method, fully embodies the advantages of computer tools applied to the study of the experiment, the experiment also shows the progress of modern means of processing computer help not only save a lot of labor time data measurement and processing, but also to ensure that the measurement accuracy of the analysis.

## REFERENCE

[1] ZHANG Chuan-wei, WANG Yun-rui, "Implementation of Serial Communication Between PLC and PC", Modular Machine tool & Automatic Manufacturing Techique, China, No.12, 2005, pp. 62-63,66.

[2] SUN Ze-wen, "Design of a Data Acquisition and Processing System Based on LabVIEW", Electrotechnics Electric,China, No.1, 2010, pp. 16-18,21.

[3] ZHU Zhan-sheng , TAN Han-dong,"3D Visualization of Electrical Data Based on Matlab Platform", Geophysical and Geochemical Exploration, China, Vol. 36, No.2, pp. 312-316.

**Liguo Zhang** graduated from Northeast Petroleum University, China, and stayed on to teach today. From teaching period he obtained master degree in Control Theory and Control Engineering from Northeast Petroleum University, China, in 2009. In 2011, he began studying motor and electrical professional doctorate in Northwestern Polytechnical University, China. His main research interests are in permanent magnet synchronous motor and control, smart instrumentation.

# Research on Spherical Panorama Mosaic Technology

Xue Leng; Song Li; Jinhua Yang
School of Photoelectric Engineering, ChangchunUniversity of Science and Technology ,Changchun, China.
Email: lengxue79@126.com

***Abstrict*-Panorama mosaic technology is an important research content in image rendering technology, it can express the complete information on the surrounding environment. As the spherical panorama with excellent visual effects, and accord with people's habits of observation, therefore be studied by many researchers. This article on the basis of the theory of understanding panorama, proposed a panorama stitching technique based on common camera,and through experiments to verify the feasibility of this method.**

***Index Terms*—spherical panorama, match, merge, mosaic**

## I. INTRODUCTION

Panoramic image stitching technology is a hot topic in the field of image processing, It has a wide range of applications in real scene mapping, Construction of virtual reality scenes,and medical Imaging synthesis. Panorama is a wide range of scenarios covering a wide viewing angle image, Panoramic camera can be used to obtain images, but it's too expensive,so most researchers using image stitching technology to obtain a panoramic image. This paper presents a method for generating panoramic images using an ordinary camera. Use an ordinary camera shoots two or more images from the same scene with an image area of overlap, They are spliced to a wide viewing angle panoramic images.

## II. METHOD

### A. *Spherical panorama stitching schematics*


Figure1. Schematic block diagram of panorama stitching

### *B. Image Acquisition*

Panoramic image acquisition can take different approaches, Using camera with a fisheye lens, You can get a panoramic image by image stitching. However, this approach shot image distortion is more serious. So this paper will use an ordinary camera to obtain images,and generate a panoramic image by stitching technique.

The first,mount the camera on a tripod, Then shoot the same level of photos by rotating the camera position, adjust the angle of the shaft on a tripod to shoot multi-angle photos. The camera looked upwards of 60 degrees, pan the camera around to shoot; Then flat camera, Then pan the camera around to shoot; Finally, the camera 60 degrees downward, pan the camera around to shoot. As long as the overlap area neighboring between two or more images to achieve pre-set standard. We can obtained a plurality of sets of photos contained the entire scene.

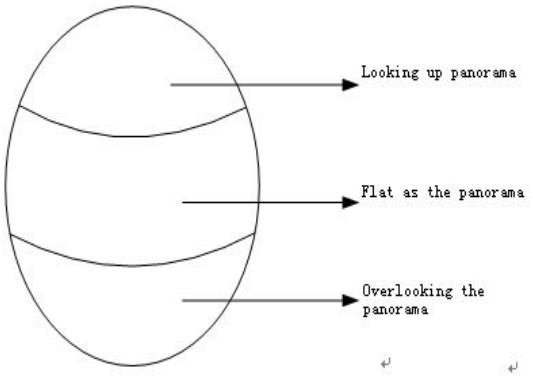



Figure2. The principle of spherical panoramas

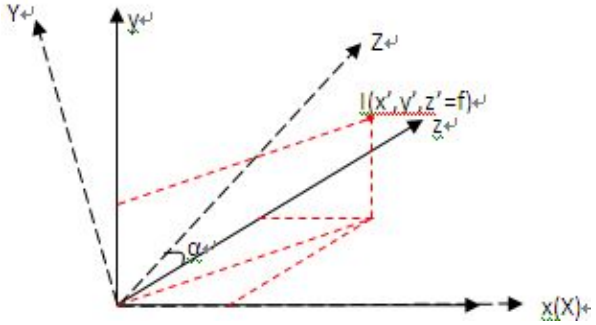



Figure3. spherical projection model

Multiple sets of the camera are captured at different angles captured, When splicing in the overlap region directly generate local deformation distortion, Therefore, the need to unify the image projected onto the sphere. After the completion of the spherical projection splicing get panoramic images without distortion. Spherical projection model used in this paper as shown in Fig3.

Let the world coordinate system is XYZ, the camera coordinate system is xyz, shooting elevation of the camera is α (for brevity, We select the camera coordinate system and the world coordinate system x (X) axis of the same shaft). Arbitrary real pixel I, the coordinates of the camera coordinate system is(x', y', f), It's coordinates in the world coordinate system is (Xw, Yw, Zw), So, the links between them can be expressed as:

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x' \\ y' \\ f \end{pmatrix} \tag{1}$$

Spherical panorama coordinates can be represented by $(\theta, \varphi, r)$ , Where $\theta$ is the horizontal angle of rotation, $\varphi$ is the pitch angle, $r$ is spherical radius, Image pixel coordinates of the spherical panorama can be expressed as:

$$\theta = tg^{-1}\frac{X_w}{Z_w} \tag{2}$$

$$\varphi = tg^{-1}\frac{Y_w}{\sqrt{X_w^2 + Z_w^2}} \tag{3}$$

The position of each image point is determined by the spherical coordinates. The following figure is an image and Its spherical projection.

Figure4. Original image

Figure5. spherical projection

*C. Image Matching*

Image matching is to find the best corresponding transformation relations from two or more images of the same scene at different times or different perspectives. There are two main image matching technique, pixel matching and feature matching. The purpose of the matching is to overcome noise and interference, looking corresponding points in two or multiple sequence images. As the feature matching method stable performance and matching speed fast, Therefore, we will use this method to complete the image matching.

In the image forming process will be affected by various factors, The image captured by the noise interference And serious distortions and aberrations. So the first thing is image preprocessing before the match completed, We use Gaussian filtering and Wallis filter combines preprocessing methods,the principleof Gaussian filter is: Using a template scan each pixel in the image, weighted average gray value determined using the template in neighborhood pixels instead of template center pixel gray value. The principle of filter Wallis is: The original image gray means and variance mapped to a given gray means and variance. After pretreatment, use harris detected corner,then corner matching for image. In the center of the corner pixels, select the gray window,the range is (2N+1)X(2N+1), use of cross-correlation method comput to reference image and the image to be matched.

$$R = \frac{\left[\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(I(x-i,y-j)-\bar{I}\right)\left(I'(x'-i,y'-j)-\bar{I}'\right)\right]}{\sqrt{\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(I(x-i,y-j)-\bar{I}\right)^2\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(I'(x'-i,y'-j)-\bar{I}'\right)^2}} \tag{4}$$

$\bar{I}$ and $\bar{I}'$ are the average of all pixels of the luminance in the gray window, get the correlation values $R$ , all corner points greater than 0.8 is regarded as a candidate, a collection of all the candidate points are not one correspondence,it has a match ambiguity. Therefore, you can calculate the strength of the candidate point matching, eliminate this ambiguity through relaxation iteration method. Specific steps are as follows:

1) Define two variables, old-total-stre and total-stre.let old-total-stre is zero, calculated the match according to the following formula, $\delta$ is0.3.

$$Stre(X_1,X_2) = \sum_{r\in M(i_1+i_2)} \frac{cor(X_1,X_2)\cdot cor(Y_1,Y_2)\cdot delta(X_1,X_2;Y_1,Y_2)}{1+\frac{\|X_1-Y_1\|+\|X_2-Y_2\|}{2}} \tag{5}$$

$$delta(X_1,X_2;Y_1,Y_2) = \begin{cases} e^{-\frac{\gamma}{\delta}} & \gamma < \delta \\ 0 & other \end{cases} \tag{6}$$

2) Define and calculate the total matching strength

$$total - strength = \sum stre(X_1,X_2) \tag{7}$$

3)Calculated at each corner of the non-ambiguity in the reference image

$$U = 1 - \frac{SecondL\arg estStre(X)}{FirstL\arg estStre(X)} \quad (8)$$

4)Compare the difference between the two variables, if less than $10^{-6}$ ,consider the overall strength and stability to match, this time exit loop to step 8, otherwise, continue to the next step.

5) All candidate matching points STRE and U, from large to small order again to Table A andB, get a new order table.

6) For the new table, Table A and B match points with the same, then that is the most similar, For a point to meet the following two conditions should be considered to eliminate the point, First, the point is in the position of the two most similar match table, Second, each point corresponding to the maximum similarity is not a the other.

7) Again iteration returns 2

8) Into an array, to match.

The results were as follows:

Figure6.Reference image

Figure7.Images to be matched

Figure8. Mosaic image

*D. Image Fusion*

Image fusion refers to an image after the completion of the match, stitching the images, and smoothing the boundary of the suture, let suture natural transition. Image fusion technology can be divided into three levels: Pixel fusion, feature fusion and decision fusion. In this paper, using the weighted average method for image fusion, it belongs to the pixel integration.LetA(x,y)is an image pixel of image A, b(X,Y)is a correspondingly pixel in image B. the fusion image isF

$$F(x,y) = \omega_1 A(x,y) + \omega_2 B(x,y) \quad (9)$$

$$\omega_1 = \frac{A(x,y)}{A(x,y) + B(x,y)} \quad (10)$$

$$\omega_2 = 1 - \omega_1 \quad (11)$$

$\omega_1, \omega_2$ are weighting coefficients for the pixel inoverlapping areas. Through reasonable choice of weighting coefficients, we can get the ideal fusion effect, achieve seamless splicing. Experimental results were as follows:

Figure9.Fusion image

## III Experiment and Analysis

Experiments done on 3.1G frequency, 2.00G memory of the PC, Experimental images shot using a tripod in a fixed viewpoint of real photos. Figure 10 is Flat as the a source image layer to be spliced, Figure 11 is a flat layer of all the pictures, as the use of the algorithm, as the layer of flat mosaic panorama.

Figure10.Fat as the a source image layer to be spliced

Figure11. Flat as thelayer panorama

Figure12. Spherical panorama

TABLE I
THE RESULTS OF THE ANALYSIS OF SPHERICAL PANORAMA STITCHING

| CAMERA TYPE | SOURCE IMAGE QUALITY | NUMBER OF IMAGE STITCHING | SPEND TIME | SUCCESS RATE |
|---|---|---|---|---|
| FUJI S205 | GOOD | 9 | 3.8SECOND | UP95 % |

## Ⅳ CONCLUSION

Experimental results show that the panorama stitching method based on the ordinary camera can eliminate traces of stitching, achieved a smooth visual effect. From the time it takes to view panorama stitching, Panorama can achieve real-time display.

## REFERENCES

[1] LIU Bo-Chao,Zhao Jian,SUN Qiang,Improved Harris Corner Detection Method Based on Edge.,Chinese Journal of Liquid Crystals and Displays,Vo1.28,PP.939-942, June 2013.

[2] Tangy Shuo, Miao Yuan,An image matching algorithm based on Harris corner, Image Processing and Multimedia Technology, Vo1.32,PP41-43, February 2013.

[3] Zhou Long Ping,Corners Detection Based on Improved HarrisAlgorithm, Computer Technology and Development, Vo1.23,PP11-14, February 2013

[4] Ruanqin,Penggang.Study on Image Registration and Mosaic TechnologyBasedon Surf Feature,Computer and Digital EngineeringPP141-144, February 2013.

[5] [Kweon Gyeong il, ChoiYoung ho,Image-processing based panoramic camera employing single fisheye lens,Journal of the Optical Society of Korea,Vol.14,pp 245-259, March 2010

[6] Brauers Johannes, Aach Til. Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras. Optics and Laser Technology,Vol.20,PP496-506.February 2012.

# BPEL-based Web Service Composition and Case Study

Lin Li
School of Computer, Wuhan University, Wuhan, China
College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
Email: lilin@wust.edu.cn

Ming Cheng[1,2]; Yong Xu[3]; GuoQing Wu[1,2]; and MengTing Yuan[1,2]
1.School of Computer, Wuhan University, Wuhan, China
2.Shenzhen Institute, Wuhan University, Shenzhen, China
3.School of Mathematics and Information Sciences, Zhaoqing University, Guangdong, China

***Abstract*—Business process execution language (BPEL) is the standard to combine Web service. It is designed to compose a group of existing services to a new Web service. Firstly, the concept of Web service composition, and the basic thought and components of BPEL were introduced; then specific description of BPEL was presented by carrying out case study, in which Web service was combined using BPEL.**

***Index Terms*—Business process execution language, Web Service Composition, Basic Activity, Structural Activity**

## I. Introduction

With the development of Internet and the characteristics of Web service, more enterprises release their services on Internet using Web service. Therefore, in order to share information among different Web services on the new platform, different services have to be combined urgently to provide a new and stronger service. Web service composition technology is proposed to solve the problem.

Owing to single Web service presents single function and fails to solve problems completely, to realize complex services, Web service has to be combined and integrated[1].

The core thought of Web service composition[2,3,4,5] is described as follow: According to users' requirements, service operations are chose automatically to meet needs by using existing Web service components. And under support of Web service composition platform, users' requests were satisfied collaboratively according to certain rules[6]. Lightweight Web service which are smaller, simpler and easier to be carried out are used to construct complex Web service which has abundant functions and is easily to be used by customers. Therefore, various relative service components which are loosely coupled and dispersed on Internet are combined to a more available system. The system enables internal and external enterprises to use the integrate Electronic Associates Incorporated (EAI) and electronic commerce.

The goal of Web service composition is to solve problems about integration and cooperation. Interactions among systems can be realized by combining Web service. So components of dispersed application are integrated. The nature of Web service composition is to perform different Web service collaboratively and flexibly satisfy more complex service requirements.

## II. Business Process Execution Language

### A. The concept of BPEL

BPEL refers to a formal specification language used in automatic business process. The processes written in BPEL in XML document can be well organized in the form of normalized interaction among W services. The processes therefore can be carried out on every platform or product in accordance with BPEL standard. So, by allowing customs to move process between kinds of authoring tools and execution platforms, BPEL saves the cost of process automation of customers. Although business process automation tends to be normalized, BPEL has attracted great attention and firstly recognized by lots of software suppliers.

### B. Basic thought of BPEL

In brief, BPEL integrates different interfaces by programming process. The aim is to provide an interface which integrates different interfaces.

First BPEL has a start point and an end point. Receive is the start point to receive its own input parameter; reply is the end point to get return parameter.

The whole BPEL is used to invoke other external interfaces and obtain the return parameter by using the input parameter of its interface.

The following definitions are most important for BPEL:

One is assignment command. It refers to assign variable values to the input parameters of interfaces which are invoked.

The other is Invoke command, which is used to call external services.

Condition command is also important. Various conditions are defined according to different variables, and then different services are invoked according to various conditions. The variables can be input parameters or the result of invoking external service. Condition

command, as an important function for running BPEL, is able to reserve and search the variables.

### C. Components of BPEL

#### a. Activities normalized by BPEL

A business task is represented as a procedure for realizing the whole business goal by activities. The standard attributes for each activity are as follows: a name, a join condition and an indicator indicating whether, or not, suppress the connection failure when it occurs.

Generally, BPEL activity is classified into basic activity and structural activity. The former is responsible for realizing certain atomic function, while the later is to integrate basic activity by using its components to realize logic function required by the system. All the activities are signed by graphical process elements and have alternative nested graphical process elements.

#### b. Partner and partner link

In BPEL, service interacted with business process is defined as partner. Each partner is described by partner LinkType. A same partner LinkType can describe different partners. Partner LinkType shows the relations between two services, defines the roles of each service, and assigns service ports for each role. The attribute value of LinkName is used for the element of source in activity.

#### c. Variable

Variable is applied to reserve information of business process. The stored information is commonly received from or going to be sent to other partners. It can be assigned as the input or output variables of activities of invoking, receiving and reply, to store data transmitted among services.

#### d. Relativity

The interaction among services is significant in business process, while loosely coupled Web service is stateless. Therefore, the concept of relativity is proposed by BPEL to solve the correlation between Web service and process case by using the characteristics of sending and receiving data.

#### e. Transaction and exception handling

A set of transaction and exception handling mechanism is provided by BPEL. By nesting a group of activities in a scope in BPEL, a transaction id produced. In the scope, failure and compensation processors can be assigned, with which, failure handling and error resilience are conducted when failures occur.

## III. CASE STUDY

### A. Problem description

A case study is carried out for the simplified business process of the arrangement of employees' business errands. Clients invoke the business process by assigning the name, the destination, the departure and the return date of employees. In BPEL business process, the employees' situation on business trip requires to be checked firstly. Assume that there is a Web service which is able to perform this service. Then, tickets prices of Eastern Airlines and Southern Airlines are surveyed in the process according to the available transportation of the employees. Finally BPEL process selects the one with lower price and returns a business trip plan to clients.

### B. Process presentation

The Web service used to detect the situation of employees on business trip is assumed to be synchronous. The data are acquired and returned to the caller immediately, so it is a reasonable method. To acquire tickets price of air, asynchronous invocation is applied. As the confirmation of flight schedule takes a long time, it is also a logic method. To simplify the case, the two airlines companies are assumed to provide Web service, and the Web services are completely same (provided with same port type and operation).

The definition of business process using BPEL actually defines a new Web service which is constituted with the existing services. The new composite Web service interface provides operations alike other Web service using a group of interface types. Therefore, business process described by BPEL has to be invoked when call for business process described with BPEL.

### C. BPEL description of process

The above arrangement of business trip is described using BPEL language as follows:

```
<process name="BusinessTripProcess"  >
 <sequence>
  <receive partnerLink="client"
portType="trv:TripApprovalPT"
operation="TripApproval"
variable="TripRequest"
createInstance="yes" />
  <assign>
   <copy>
    <from variable="TripRequest"
     part="employee"/>
    <to variable="EmployeeTripStatusRequest"
part="employee"/>
   </copy>
  </assign>
  <invoke  partnerLink="employeeTripStatus"
portType="emp:EmployeeTripStatusPT"
operation="EmployeeTripStatus"
inputVariable="EmployeeTripStatusRequest"
outputVariable="EmployeeTripStatusResponse" />
  <assign>
   <copy>
    <from variable="TripRequest" part="flightData"/>
    <to variable="FlightDetails" part="flightData"/>
   </copy>
   <copy>
    <from variable="EmployeeTripStatusResponse"
part="TripClass"/>
    <to variable="FlightDetails" part="TripClass"/>
```

```
    </copy>
   </assign>
   <switch>
<case condition="bpws:getVariableData('FlightResponse
East', 'confirmationData','/confirmationData/Price')
<= bpws:getVariableData('FlightResponseSouth',
'confirmationData','/confirmationData/Price')">
   <assign>
    <copy>
     <from variable="FlightResponseEast" />
     <to variable="TripResponse" />
    </copy>
   </assign>
   </case>
   <otherwise>
   <assign>
    <copy>
     <from variable="FlightResponseSouth" />
     <to variable="TripResponse" />
    </copy>
   </assign>
   </otherwise>
   </switch>
   <invoke partnerLink="client"
portType="trv:ClientCallbackPT"
operation="ClientCallback"
inputVariable="TripResponse" />
  </sequence>
</process>
```

## IV. CONCLUSION

Business Process Execution language BPEL integrates Web service, which is supplied by different service providers and uses different technologies on different platforms. Therefore, it provides significant functions for users in the form of function body. The basic thought, structure and characteristics of BPEL were analyzed in this paper; additionally, by analyzing a specific case of business trip, the process of combining Web service using BPEL and the specific BPEL description were obtained. The process showed certain practical value.

## ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 91118003,61003071,61100055) Special Funds for Shenzhen Strategic New Industry Development (JCY20120616135936123), Wuhan University of Science and Technology Youth Training Plan(2014xz017).

## REFERENCES

[1] Fu Yanning, Research on Web service composition methods (Ph.D. Thesis). Changchun: Jilin University, 2007.
[2] Yue Kun, Wang Xiaoling, Zhou Aoying.Web service core support technology: research overview. Journal of software. 2004,15(3):428-442.
[3] Rao Yun, Feng Boqin, Li Zhunchao. Overview of Web services composition based on Services. Systems engineering and electronics, 2005,27 (8): 1481-1489.
[4] Biplav Srivastava, Jana Koehler. Web Service Composition: Current Solutions and Open Problems. In: Proc. of the 13th International Conference on Automated Planning & Scheduling(ICAPS2003),Workshop on Planning for Web Services, 2003:28-35.
[5] Schahram Dustdar, Wolfgang Schreiner. A survey on web service composition. Int. J. Web and Grid Services,2005,l(l):l-30.

# Empirical Study of dynamic Endogeneity: A Panel Vector Regression Approach

Lin Shao
School of Finance, QiLu University of Technology, Jinan, China
Email: shaolinweihai@163.com

Byunghua Lim
School of economics and managment , Suwon University, Suwon,Korea
Email: byunghwalin@suwon.ac.kr

***Abstract*—This study offers new insights into the dynamic relationship between ownership structure and firm performance in the case of Chinese Stock Exchange in the period of 1999 to 2012 using a panel Vector Auto-regression (PVAR). The panel VAR approach allows to addresses the endogenous problem by allowing the endogenous interaction between the variables in the system. Our result provides the evidence that dynamically inter-temporal relationship exists between ownership and performance for a long run. Impulse response and variance decomposition illustrate that impact of a shock to variables themselves is the main source for their variability.**

***Index Terms*—ownership structure, performance, dynamic endogeniety, panel VAR**

## I. Introduction

A considerable amount of empirical researches have verified that dynamic endogenous relationship exists in the corporate governance. Hu and Izumida (2008) argue that ownership concentration has a significant effect on contemporary and subsequent corporate performance. Fahlenbrach and Stulz (2009) find that the relationship between change of performance (Tobin'Q) and past and contemporaneous change in ownership structure depend on controlling for past stock returns. Cheung and Wei (2006) also indicate that insider ownership and corporate performance can be explained by their respective lagged values. Wintoki et al (2012) summarize that there are three main sources of endogenous problem in empirical experiments, including unobservable heterogeneity, simultaneity and current values of governance variables are a function of past firm performance. The third endogeneity is called as dynamic endogeneity. In their study, there is no relation between current board structure and current firm performance when taking account of dynamic endogeneity. Davidson and Rowe (2004) also develop a theory of inter-temporal endogeneity of board composition and financial performance. Using causality tests in panel regressions with three years of data for 130 closed-end mutual funds, they find that only minimal evidence of inter-temporal endogeneity.

Prior relevant empirical researches have shown that dynamic endogeneity exists in the corporate governance and neglecting this endogenous problem can have serious consequences for inference.

Regarding to the methodology of exploring the dynamically endogenous problem, Davidson and Rowe (2004) use fixed effect model and random effect model to exploit the endogenity between board composition and financial performance. Wintoki et al (2012) utilize the dynamic panel generalized method of moments (GMM) to investigate the relation between board structure and firm performance. Different from the aforementioned method, this study provides a pioneering approach by applying panel vector autoregression method (PVAR) to examine the possible endogenous correlation between ownership and performance initially.

The idea that ownership and performance may be endogenously related is not new. However, Panel VAR method treats all variables as endogenous variables in a system and uses orthogonalized impulse-response functions, which shows the responses of one variable of interest (i.e. ownership) to an orthogonal shock in another variable of interest (i.e. performance). By orthogonalizing the response we are able to identify the effect of one shock at a time, while holding other shocks constant (Love, 2006). Panel VAR approach is a method of system, which is similar to prior simultaneous equation system. Panel VAR mainly explore the dynamic relation between different variables in a system.This paper attempts to look for some new findings using panel VAR.

Sims (1980) provides a new macro-econometric framework: vector auto-regressions (VARs). A univariate auto-regression is a single-equation, single-variable linear model in which the current value of a variable is explained by its own lagged values. A VAR is an *n*-equation, *n*-variable linear model in which each variable is in turn explained by its own lagged values, plus current and past values of the remaining *n-1* variables. This simple framework provides a systematic way to capture rich dynamics in multiple time series and the statistical toolkit that came with VARs was easy to use and interpret. As Sims (1980) and others argued in a series of influential early papers, VARs held out the

promise of providing a coherent and credible approach to data description, forecasting, structural inference, and policy analysis.

In this study, a dynamic model like a VAR is an appropriate choice for estimating the interrelationship between ownership structure and firm performance as a dynamic process.

Observing previous studies of ownership-performance, no vector auto-regressions model is to consider the dynamic interactive impacts between ownership structure and firm performance. Dynamic panel generalized method of moment (GMM) is also a main method to alleviate potentially dynamic endogeneity comparing with traditional methods such as ordinary least squares, fixed effect model and simultaneous equation model. In addition, a considerable amount of prior studies have ignored dynamic analysis solution, such as generalized method of moment, impulse responses and variance decompositions, and have had gaps in their econometric procedure of applying the VAR model, such as ignoring VAR diagnostics. All of these factors may have caused biased results. The paucity of literature encourages me to engage in further study, but with a different approach to correct current shortcomings. In an attempt to decompose cause and effect, we estimate panel vector auto-regressions (PVAR) generated by GMM that describe the dynamic relation between ownership structure and firm performance.

There are several possible contributions to a growing number of recent literatures of corporate finance. Specifically, this paper is in line with previously cited studies on the endogenous interaction between ownership and firm performance in corporate conditions.

Firstly, the study provides the econometric application to avoid misspecification and to minimize the resulting bias. It tests and estimates the causal relationship by applying the three-variable VAR model based on the panel data (ownership structure, investment and firm performance). We use vector auto-regressions on panel data and enable us to investigate the endogenous interrelationship between ownership structure and firm performance in corporate finance, which allowing for a firm-specific unobserved heterogeneity considering the levels of variables (i.e. fixed effects). Most of past literatures normally use dynamic panel data model to deal with the correlation between ownership and performance and use the lag one term of dependent variable, we wish we can provide some new results by using vector auto-regressions.

Secondly, using the orthogonalized impulse-response functions, we are able to separate the response of one variable to shocks coming from other variables. Traditional methods (fixed effect model or simultaneous equation) are not able to complete this work.

This paper also supplements the scant literature on relationships between ownership structure and firm performance and empirical evidence about the source of endogeneity by using a new approach: panel VAR. This paper also adds to present the new evidence or of corporate governance that ownership structure and firm performance in China from 1999-2012.

This paper also analyzes and maps economic policy onto estimated results, and then provides insightful policy implications for governments.

The purpose of this section is to investigate the dynamic interrelationship between ownership structure and firm performance. The main objective is to study whether the dynamics of firm performance can be used as indication of the change of ownership structure or not. To test this hypothesis, investment variable is involved in the panel VAR system. For doing this, we document significant differences in the response of investment to firm performance because McConnel and Muscarella (1985), Cho (1998) have shown that investment positively affects corporate value. Furthermore, we test the response between ownership structure, investment and firm performance and assume them endogenously.

The rest of the paper is as follows: section 2 presents the empirical methodology specification including data description, unit root test and lag length selection; Section 3 provides the results and summary and conclusion is presented in section 4.

## II. EMPIRICAL METHODOLOGY

### *A. Sample and Data*

The sample utilized in this study comprises data for 350 public companies listed on the Chinese Stock Exchange quoted on the Shanghai and Shenzhen. Annual dataset was collected for these companies in respect of the period 1999 to 2012 inclusive. The total effective number of observations is 4900. Data predominantly was obtained from three sources: The first database is the Chinese Center for Economic Research (CCER).The second database is the China Stock Market and Accounting Research (CSMAR) database. The third database is RESSET database.

The sample of firms employed in the study is subject to the following criteria: firstly, remove unavailable information, indeterminable ownership structure and incomplete financial data; secondly, eliminate companies treated by ST, *ST and PT; thirdly, exclude firms of issuing both B and H shares;lastly, excluded financial companies;(4) the firm must have been quoted on the Chinese Stock Exchange at least 1 year before year of analysis.

In our investigation, we use a panel-data vector auto-regression methodology with three variable, including ownership, investment and performance. All variables are treated as endogenous variables in our study, the interrelationship between these variables can be tested efficiently.

Ownership concentration is measured by the fraction of share owned by the first largest shareholder (CR). Investment is calculated by net capital expenditure divided by the total assets (CAPITAL). Performance variable is proxies for two alternative variables: return on assets (ROA) and Tobin'Q (Q). ROA variable is utilized

to test the robustness of variable sensitivity. Table I reports the summary statistics for the firm-level variables.

TABLE Ⅰ.
STATISTICS DESCRIPITON

| variable | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|
| Q | 2.08 | 1.31 | 0.59 | 1.69 | 14.98 |
| ROA | 0.04 | 0.07 | -0.97 | 0.03 | 2.68 |
| CR | 0.40 | 0.17 | 0.04 | 0.38 | 0.89 |
| CAPITAL | 0.07 | 0.09 | -0.91 | 0.04 | 1.48 |

*B. Unit Root Test*

Unit root test is a necessarily initial step for estimation using panel VAR model. In this study we implement two panel unit root tests (LLC and ADF tests) proposed by Levin et al. (2002), Maddala and Wu (1999), respectively. The null hypothesis of the above unit root tests is that there exist unit root in the series, i.e., the variables are non-stationary. Rejecting the null hypothesis means the series is stationary. This series is non-stationary if we cannot reject the null hypothesis. Unit root test is reported in Table Ⅲ.

TABLE Ⅱ.
UNIT ROOT TEST

| statistics | Levin-Lin-Chu | | ADF Fisher | |
|---|---|---|---|---|
| | Trend | No trend | trend | No trend |
| Variable | Adjusted t*Statistics | Adjusted t*Statistics | Chi-squared Statistics | Chi-squared Statistics |
| ROA | -21.88*** | -18.74*** | 15.31*** | 23.29*** |
| Tobin'Q | -30.41*** | -44.17*** | 14.76*** | 47.46*** |
| CR | -61.73*** | -25.31*** | 4.89*** | 1.96** |
| CAPITAL | -28.75*** | -31.71*** | 25.26*** | 32.88*** |

Notes : *** indicates significance at 1% level, ** at 5% and *** at 10% level, respectively.

*C. Model Specification*

We mainly use a panel-data vector auto-regression methodology. This technique contains the traditional VAR approach, which treats all the variables in the system as endogenous, with the panel-data approach, which allows for unobserved individual heterogeneity.

To implement the model, the empirical methodology, which is closely based on the approach taken in Love and Zicchino (2006), is to estimate a P-order ***n***-variable VAR model in a panel setting as follows:

$$y_{it} = \alpha_0 + \alpha_1 y_{i,t-j} + f_i + d_t + \varepsilon_{it} \quad (1)$$

Where $y_{it}$ is one vector of endogenous variable. The vector has three variables containing CAPITAL, Q and CR. $\alpha_1$ is a vector of parameters to be estimated, $f_i$ represents firm -fixed effects. $d_t$ denotes time effects, $\varepsilon_{it}$ is the error term assumed to be IID with a zero mean. The lowercase subscripts *i* and *t* represent firm *i* at time *t* respectively, with the period t (1999-2012). The VAR includes $j$ lags, which is selected using the Information Criterion.

In applying the VAR procedure to panel data, we need to impose the restriction that the underlying structure is the same for each cross sectional unit. Since this constraint is likely to be violated in practice, one way to overcome the restriction on parameters is to allow for "individual heterogeneity" in the levels of the variables by introducing fixed effects, denoted by *i* in the model (Love and Zicchino, 2006). Since the fixed effects are correlated with the regressors due to lags of the dependent variables, the mean-differencing procedure commonly used to eliminate fixed effects would create biased coefficients. To avoid this problem we use forward mean-differencing, also referred to as the 'Helmert procedure" (Arellano and Bover, 1995). This procedure removes only the forward mean, i.e. the mean of all the future observations available for each firm-year. This transformation preserves the orthogonality between transformed variables and lagged regressors, so we can use lagged regressors as instruments and estimate the coefficients by system GMM method. This is a standard procedure for estimating dynamic models with panel data.

*D. Lag length selection*

This step is to check the lag order selection. The model will be over-parameterized if the number of lags is too large. Too-long lags result in a rapid loss of degrees of freedom and over-parameterization, while too-short lags might introduce biased results caused by omitting important variables and failing to capture the system's dynamics.

In our study, AIC, SC and HOIC are simultaneously selected as the criterion of the lag order selection. The appropriate lag length for panel VAR model is one is presented in Table Ⅲ.

TABLE Ⅲ.
SELECTION ORDER CRITERION

| *lag* | *AIC* | *BIC* | *HQIC* |
|---|---|---|---|
| 1 | -2.51049 | -.911184 | -1.94498 |
| 2 | -2.6978* | -.962418* | -2.08146* |
| 3 | -2.62208 | -.726403 | -1.94554 |
| 4 | -2.50365 | -.416063 | -1.75461 |
| 5 | -2.33226 | -.010323 | -1.49413 |
| 6 | -1.98609 | .629196 | -1.03564 |
| 7 | -1.66478 | 1.32955 | -.568039 |
| **8** | -1.29661 | 2.20875 | -.000807 |

Note:(*) indicates lag order selected by the criterion
AIC: Akaike information criterion
SC: Schwarze information criterion
HQIC: Hannan-Quinn information criterion

## Ⅲ. EMPIRICAL RESULT

This section analyzes the primary result through panel VAR model of three-variable panel VAR model by applying GMM model. The lag orders chosen by the AIC criterion, HQIC criterion and the BIC criterion are two. Table Ⅳ provides the primary result of panel VAR model. Impulse response is reported in Figure 1. Variance Decomposition is presented in Table Ⅴ.

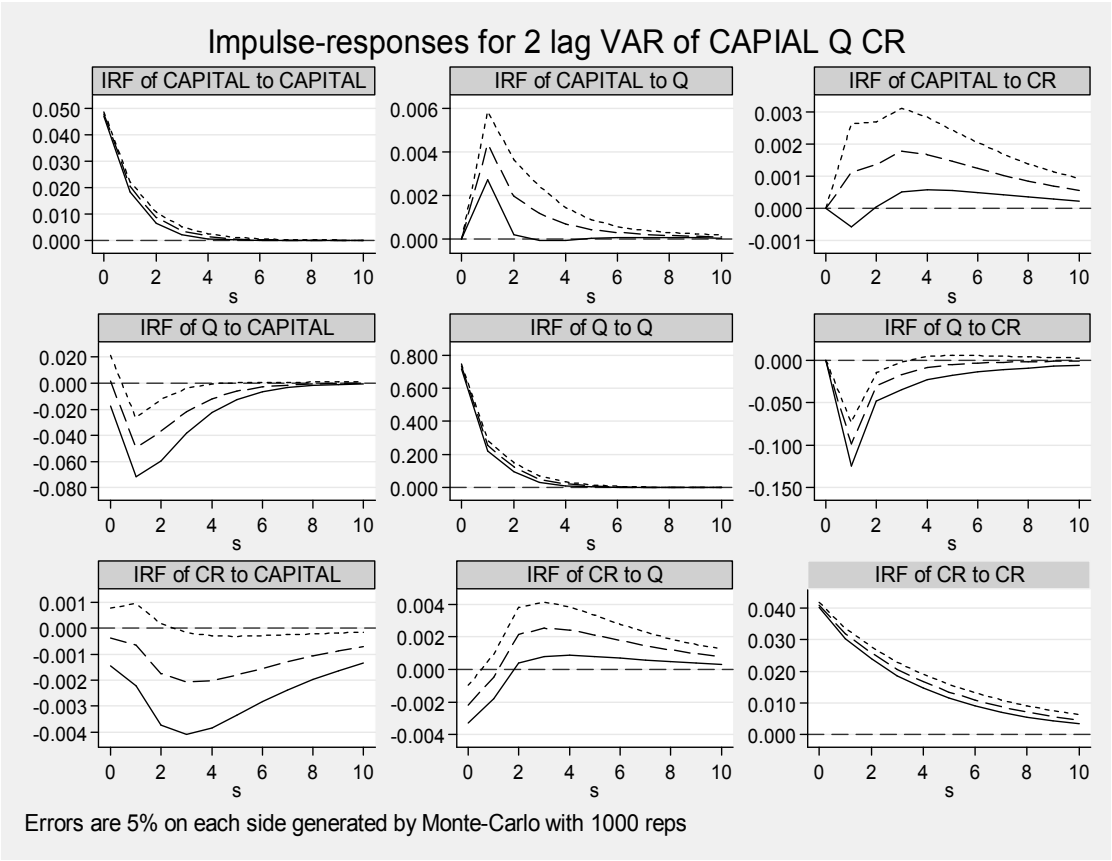



Figure 1.Impulse-responses for 2 lag VAR of CAPITAL Q CR

*A. Result of panel VAR*

TABLE Ⅳ.
RESULT OF PVAR MODEL

| | CAPITAL | Q | CR |
|---|---|---|---|
| L.CAPITAL | 0.426*** | 1.057*** | -0.007 |
| | (18.24) | (-4.02) | (-0.46) |
| L.Q | 0.006*** | 0.339*** | 0.002* |
| | (4.42) | (13.02) | (1.91) |
| L.CR | 0.027 | -2.393*** | 0.778*** |
| | (1.17) | (-6.23) | (35.10) |
| L2.CAPITAL | 0.005 | 0.008 | -0.021 |
| | (0.26) | (0.03) | (-1.52) |
| L2.Q | -0.002* | 0.062*** | 0.003*** |
| | (-1.78) | (3.47) | (4.03) |
| L2.CR | 0.015 | 1.965*** | 0.030** |
| | (0.91) | (5.93) | (1.99) |

Note: Heteroskedasticity adjusted t-statistics are in parentheses. *** indicates significance at 1% level, ** at 5% and *** at 10% level, respectively. L.CR is the first order difference of CR..

*B. Result of impulse response*

Result of panel VAR model using GMM method indicates that there is dynamically inter-temporal relationship between ownership and performance. It is interested noting that there is significantly negative relation between current Q and lagged one CR. However, the coefficient of lagged two CR is positive and statistically significant. Impulse responses also illustrates that the response of Q to CR shock is negative from the beginning to the second period, and then shows a positive shock after the second period. We also can find that there is a long interactive impact between CR and Q, and the persistent period is ten years or more.

*C. Result of varance decomposition*

Table Ⅴ
VARIANCE DECOMPOSITION

| | PERIOD | CAPITAL | Q | CR |
|---|---|---|---|---|
| CAPITAL | 1 | 1.000 | 0.000 | 0.000 |
| CAPITAL | 2 | 0.993 | 0.007 | 0.000 |
| CAPITAL | 3 | 0.991 | 0.008 | 0.001 |
| CAPITAL | 4 | 0.989 | 0.009 | 0.002 |
| CAPITAL | 5 | 0.988 | 0.009 | 0.003 |
| Q | 1 | 0.000 | 1.000 | 0.000 |
| Q | 2 | 0.004 | 0.980 | 0.016 |

| | | | | |
|---|---|---|---|---|
| Q | 3 | 0.006 | 0.977 | 0.017 |
| Q | 4 | 0.007 | 0.976 | 0.017 |
| Q | 5 | 0.007 | 0.976 | 0.017 |
| CR | 1 | 0.000 | 0.003 | 0.997 |
| CR | 2 | 0.000 | 0.002 | 0.998 |
| CR | 3 | 0.001 | 0.003 | 0.996 |
| CR | 4 | 0.002 | 0.004 | 0.994 |
| CR | 5 | 0.003 | 0.005 | 0.992 |

Variance decomposition indicates that the fluctuations of Q are explained mainly by Q shocks and CR shock in a long run. CR shock accounts for 100% in the first and second year from CR shock and Q shock. As for Q, the fluctuation of Q is dominantly explained by itself. CAPITAL shock just plays a little role in explaining the effect of its shock to ROA and Q. CR is only explained by itself in a long run. In summary, shocks to CR, CAPITAL and ROA are important sources of variability for themselves.

### *D. Robustnes test*

In this section, we conduct several additional tests to investigate the sensitivity of our results, which are not reported here in the interest of brevity. We introduce the ROA into the panel VAR model to test the stability of data. The evidence indicates that the result maintain the expected signs.

## Ⅳ. CONCLUSIONS

This paper uses a VAR approach to analyze the relationship between ownership and performance. It shows that there indeed is a dynamically endogenous relationship between ownership and performance for a long run. More specifically, the impact of a shock to ownership (Q) on ownership (CR) is positive and continues for a long period. The impulse response of performance to ownership shock is a U shaped relation, namely, negative impact is in the beginning and then changes positive impact in the next periods.

## REFERENCES

[1] Yabei Hu and Shigemi Izumida, "Ownership concentration and corporate performance: A causal analysis with Japanese panel data," Corporate Governance, vol. 16, July 2008.

[2] M. Babajide Wintoki, James S. Linck, Jeffry M. Netter, "Endogeneity and the dynamics of internal corporate governace," Journal of Financial Economics, vol. 105, pp.581–606, 2012.

[3] W. K. Adrian Cheung, K. C. John Wei, "Insider ownership and corporate performance: Evidence from the adjustment cost approach," Journal of Corporate Finance, vol. 12(5), pp. 271–350, 2006.

[4] Wallace N. Davidson, Wei Rowe, "Intertemporal endogeneity in board composition and financial performance," Corporate ownership & Control, Vol. 1(4), 2004.

[5] Rudiger Fahlenbrach, Rene M. Stulz, "Managerial ownership dynamics and firm value," Journal of Financial Economics, Vol. 92, pp.342-361, 2009.

[6] McConnell, J., Muscarella, C., "Corporate capital expenditure decisions and the market value of the firm," Journal of Financial Economics, Vol. 14, pp.399-422, 1985.

[7] Inessa Love, Lea Zicchino, "Financial development and dynamic investment behavior: Evidence from panel VAR," The Quarterly Review of Economic and Finance, Vol. 46, pp.190-210, 2006.

[8] Cho, M., "Ownership structure, investment and the corporate value: An empirical analysis," Journal of Financial Economics, Vol. 47, pp.103-121, 1998.

[9] Levine A， Lin C F, Chu C S, "Unit root tests in panel data: Asymptotic and finite sample properties," Journal of Economics, Vol. 57, pp.265-301, 2002.

[10] Madala G S and Wu, "A comparative study of unit root tests with panel data and a new simple test," Oxford Bulletin of Economics and Statistics, Vol. 61, pp. 631-652, 1999.

[11] Arellano, M., Bover, O., "Another look at the instrumental variable estimation of error component models," Journal of Economics, Vol. 68, pp.29-51, 1995.

[12] Sims, C. "Macroeconomics and Reality," Econometrica, Vol.48, pp.1-48, 1980.

[13] Demsetz, H., Villalonga, B., "Ownership structure and corporate performance,"Journal of Corporate Finance, Vol.7. pp.209-233, 2001.

[14] Love,I. "Financial development and financing constraints: International evidence from the structural investment model," Review of Financial Studies, Vol.16, pp.765-791, 2003.

[15] Himmelberg, C. P., Hubbard, R.G., Palia, D.," Understanding the determinants of managerial ownership and the link between owenrship and performance, " Journal of Financial Economics, Vol. 53, pp.353-384, 1999

# The Wheat Prices Prediction based on ARIMA Model

Shiliang Du
School of Information Science and Engineering, Henan University of Technology, Zhengzhou, China
Email:dsl1989@126.com

Dexian Zhang
School of Information Science and Engineering, Henan University of Technology, Zhengzhou, China

***Abstract*—China is a big country of wheat production, and the wheat prices is related to the fundamental interests of the majority of farmers, many researchers have been studying on predicting the wheat prices. However, there are many factors affecting the price, so their study did not accurately predict the price of wheat. In this paper, we used the method of time series for predicting the wheat prices and selected the price data from January 2010 to February 2014, through pattern recognition, establishing Autoregressive Integrated Moving Average Model (ARIMA) models to predict China wheat market from March 2014 to June of wheat prices. By comparing data with the actual value, we determine the permissible error range and eventually give four-month trend of wheat prices from July 2014 to October 2014, which Can provide theoretical support for the development of national food policy in accordance with predictable results.**

***Index Terms*—Wheat Price, Prediction, Spss, ARIMA**

## I. INTRODUCTION

China is a large population and agricultural country, with the world's largest wheat production, and wheat is as a major rations for chinese people, it has close relationship with national security and social stability. Wheat supply and demand balance has been the focus of government attention. Therefore, with the current economic situation, we use appropriate mathematical model to make quantitative predictions of the wheat market price. Under this background, the study of China wheat market price prediction is of great significance, which can provide realistic theory support for China wheat market development and policy setting, and promote Chinese wheat industrial management and the agricultural modernization, so as to stabilize farmers' incomes and safeguarding national food security[1].

At present, in terms of price prediction the H-P filter analysis is used. Its basic idea is that separating long-term trend component and a component of short-term fluctuations from economic fluctuations. In the cycle component of the standard deviation, and then filtering or Fourier analysis of variance (Fourler) filtering method for filtering cycle analysis, the remaining part of the random fluctuations, operating rules by analyzing the various waveforms, judge and predict a time series. There are other commonly used correlation analysis, cointegration, Granger causality test method, stationary time series and non-stationary time series methods.

To some extent, Selecting the prediction method determines the accuracy of the predicted results[2]. Because China wheat prices is subject to a variety of factors, and complex relationship exists between factors and extremely ,Therefore, by using of structural causal model is to predict china wheat prices, it is generally difficult to achieve good prediction[3][4][5].

Time series analysis techniques, after decades of development, has become a relatively complete and independent discipline, it is used in past behavior sequence build modeling to analyze the impact of the current, without repeating considering other equences factors affection, namely the goal changes predicted of all the factors affecting the by all the "time" and are described together. ARIMA (Autoregressive Integrated Moving Average Model) model is based on variable variation itself, and the use of time-series change extrapolation mechanism is described, focusing on analysis of probability or random nature of economic time series itself, regardless of the economic theory as the basis of the explanatory variables role, it's for non-stationary time series prediction[6].

This article aimed at regardless of other factors, by the method of time series using ARIMA models to predicted future price trend in wheat market situation. Since the wheat market prices sometimes vary widely, by selecting the monthly average data can ignore this effect, and finally gives an appropriate predictive model.

## II. ARIMA MODEL

ARIMA model is proposed statisticians Box and Jenkins, also known as BJ model. It depends principle is: Some time series is dependent on the time t of a set of random variables, although the values constitute a single sequence is uncertain, but there are some changes in the sequence of regularity, we can use the corresponding the mathematical model (ARIMA) approximate description[7][8]. According to advance research on the characteristics of time series, we used three parameters to analyze time series, namely autoregressive order (p), the differential frequency (d) and the moving average order (q), usually the model is written as ARIMA (p , d, q)[9][10].

There are four basic types of ARIMA models: autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) models and differential autoregressive moving average model (ARIMA).

AR model believes that the same form since the time series regression model with the general linear regression model, the only difference is that the explanatory variables in the model are the explanatory variables 1,2,3, ..., p-order lag variable. Therefore, p-order autoregressive model AR (p) can be used in equation (1) is expressed as:

AR model believes that the same form since the time series regression model with the general linear regression model, the only difference is that the explanatory variables in the model are the explanatory variables 1,2,3, ..., p-order lag variable. Therefore, p-order autoregressive model AR (p) can be used in equation (1) is expressed as:

$$u_t = c + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \cdots + \alpha_p u_{t-p} + \beta_t \quad (1)$$

Among them: the parameter c is a constant; parameters $\alpha_1, \alpha_2, \cdots, \alpha_q$ is the p-order moving average model coefficients; parameter $\beta_t$ mean of zero and variance $\delta^2$ of the white noise sequence.

From AR (p) is defined, after the removal of indirect relevance, with its sequence of intervals over p values will no longer be relevant, and therefore AR (p) of the partial autocorrelation function p-order function after presenting censored. Requirements for stationary AR(p) model, through $\alpha_1, \alpha_2, \cdots, \alpha_q$ reflected.

MA model believes that the time series model can be based on the principle of the average, prediction error of the pre-established predictive value over the previous period plus a prediction error can get now predicted values. So, after the recursive MA (q) model can be expressed by equation (2):

$$u_t = \mu + \beta_t + \theta_1 \beta_{t-1} + \theta_2 \beta_{t-2} + \cdots \theta_q \beta_{t-q} \quad (2)$$

Among them: parameter $\mu$ is a constant; parameter $\theta_1, \theta_2 \cdots \theta_q$ is q-order coefficient moving average models; $\beta_t$ are zero mean and variance $\delta^2$ of the white noise sequence.

From MA (q) is defined: the moving average model $\beta_t$ is a linear combination of q+1 a recent value of white noise sequence, therefore, $\beta_t$ would only affect the value of q + 1 of the sequence $u_t$.

ARMA (p, q) model is built on the AR (p) and MA (q) model based on. It can be expressed by equation (3):

$$u_t = c + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \cdots + \alpha_p u_{t-p} + \beta_t + \theta_1 \beta_{t-1} + \theta_2 \beta_{t-2} + \cdots + \theta_q \beta_{t-q} \quad (3)$$

Wherein p,q, respectively highest order different from zero, from the partial correlation function values and autocorrelation function values are significantly. ARMA (p, q) models are available for sequence fewer parameters to better matching, their autocorrelation and partial autocorrelation function showed a tailing.

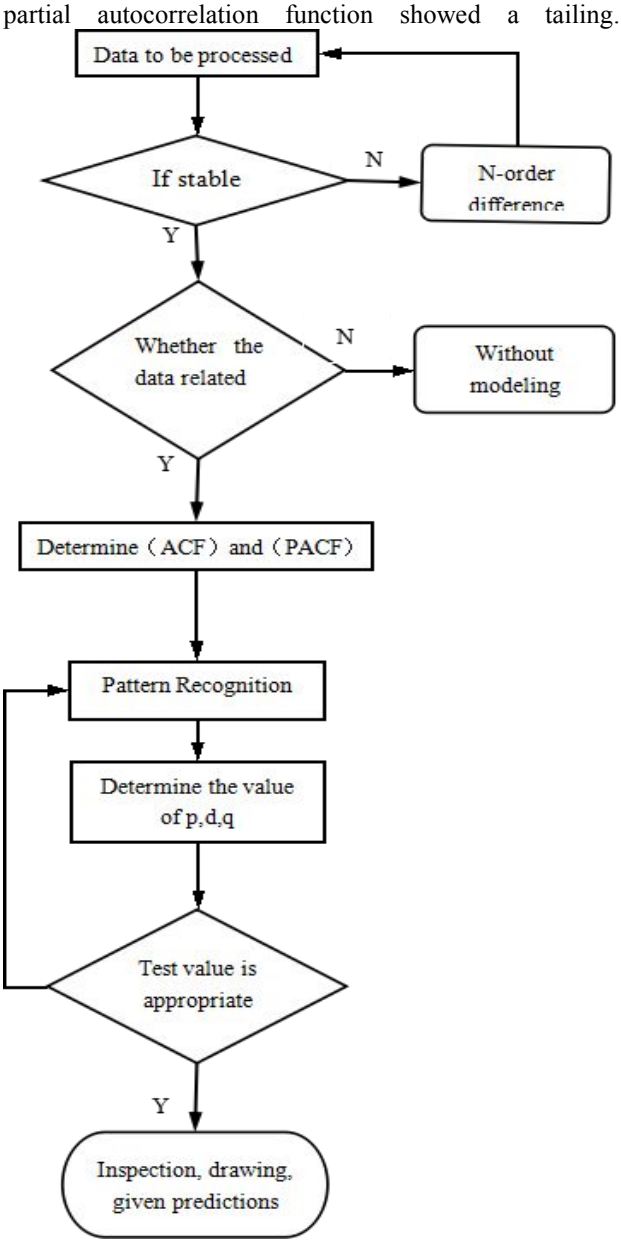



Fig.1 ARIMA modeling steps

Since the ARMA model is only suitable for stationary series analysis, for non-stationary time series if built directly ARMA model often appear spurious regression problem. However, in the practical application of time series are not stationary series, it is not directly ARMA model, but usually these sequences can be processed into a d-order difference stationary series, which is used in this paper ARIMA (p, d, q) models.

## III. ESTABLISH AND PREDICTION ANALYSIS OF ARIMA MODEL

The general steps ARIMA time series analysis and forecasting model is shown in figure(1).

This paper monthly 54 groups average data as sample data shown in Table (1), the data source for Chinese food

network (www.cngrain.com) nationwide average price of wheat historical data. Sampling period of from January 2010 to February 2014, a total of 50 sets of data use spss software simulation process.

TABLE 1
THE NATIONAL AVERAGE PRICE OF WHEAT(YUAN·TON-1)

| | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|
| Jan | 2006 | 2087 | 2109 | 2509 | 2568 |
| Feb | 2000 | 2104 | 2113 | 2512 | 2572 |
| Mar | 2008 | 2126 | 2117 | 2501 | 2567 |
| Apr | 2010 | 2115 | 2140 | 2480 | 2518 |
| May | 2004 | 2098 | 2146 | 2482 | 2505 |
| Jun | 1930 | 2030 | 2095 | 2388 | 2432 |
| Jul | 1935 | 2016 | 2068 | 2416 | |
| Aug | 1953 | 2017 | 2116 | 2441 | |
| Sep | 1970 | 2026 | 2205 | 2468 | |
| Oct | 2024 | 2102 | 2245 | 2536 | |
| Nov | 2095 | 2106 | 2305 | 2571 | |
| Dec | 2080 | 2111 | 2362 | 2573 | |

Between 2010-2014 in February wheat prices time series shown in figure (2), Preliminary judgment is not smooth by the timing diagram, Therefore, use the sequence analysis unit root testas shown in figure(3),we can know that the w heat prices time series of non-stationary time series. The sequences do get the differential second- order difference time series shown in figure (4). As can be seen, the results of the differential after its up and down around zero

Baseline fluctuations, At this point we can assume that it has reached a plateau, for if really smooth, we did the unit root test. Unit root test results are shown in figure(5), can be seen through the sequence after the second-order differential t-test statistic is less than the critical value for each given level of significance, which can be judged, the time series after the second order differential series is stationary sequence.

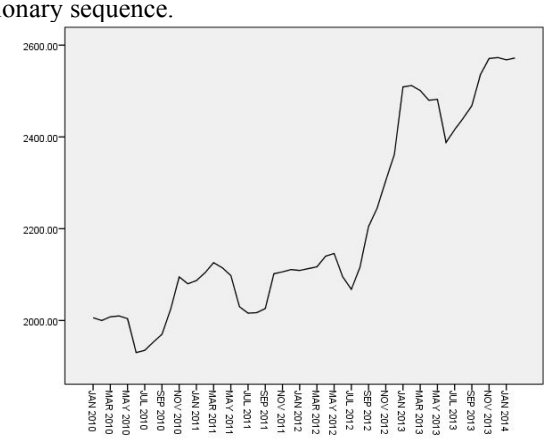



Fig.2 Original time sequence diagram

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -1.185050 | 0.6743 |
| Test critical values: | 1% level | -3.562669 | |
| | 5% level | -2.918778 | |
| | 10% level | -2.597285 | |

Fig.3 Second order differential time series

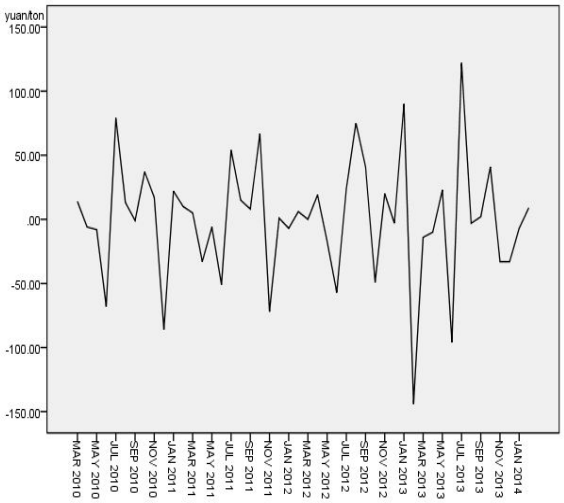



Fig.4 Original sequence of unit root test results

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -9.976423 | 0.0000 |
| Test critical values: | 1% level | -3.565430 | |
| | 5% level | -2.919952 | |
| | 10% level | -2.597905 | |

Fig.5 Original sequence of unit root test results

The second-order difference Time series is a stationary time series, thus can determine parameter d values of the ARIMA (p, d, q) is 2. Through analysis the second-order differential autocorrelation sequence (Figure 6) and partial correlation (Figure 7) ,we know that the sequence relevant tail after four relevant. partial autocorrelation tail after one relevant. Based on the ARIMA model determined p, q values way , p may take 1, q can take 1-4. Analysis of AIC and BIC results under different p q values, obtained p take 1, q take 1 is an optimal model. That is the price of wheat is prediction model is the ARIMA (1,2,1).

According determine the optimal model ARIMA (1,2,1) predict March 2014 to October. the fitting results of the resulting image shown in figure(8) and table (2).

As can be seen from Table (2), the predicted value and the actual value little difference, the average prediction error from March to June of only 0.8075%, indicating that ARIMA (1,2,1) model to predict short-term price of wheat is better . Therefore, ARIMA (1,2,1) model can be used predict wheat price.

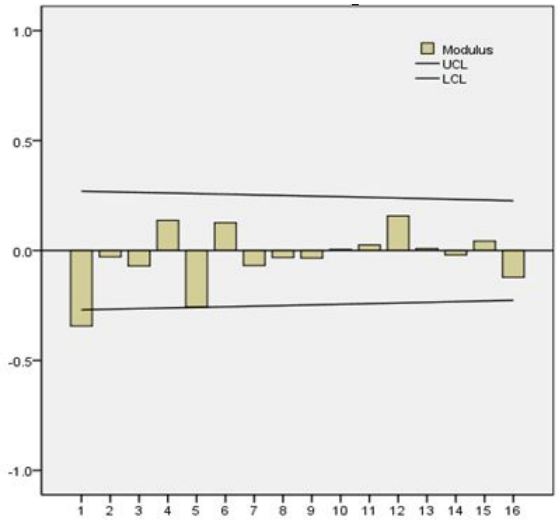



Fig.6 Autocorrelogram

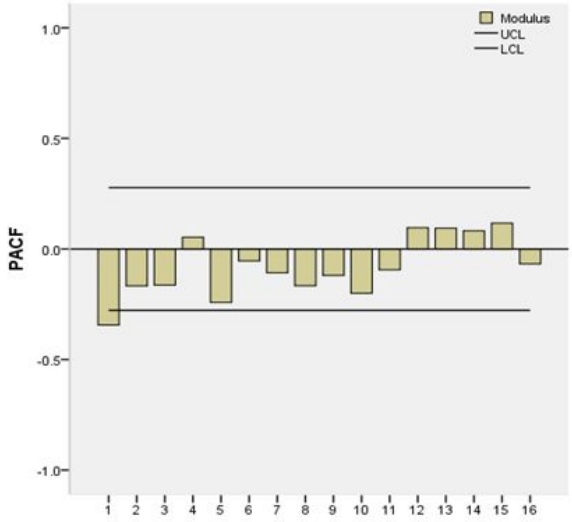



Fig.7 Partial autocorrelation

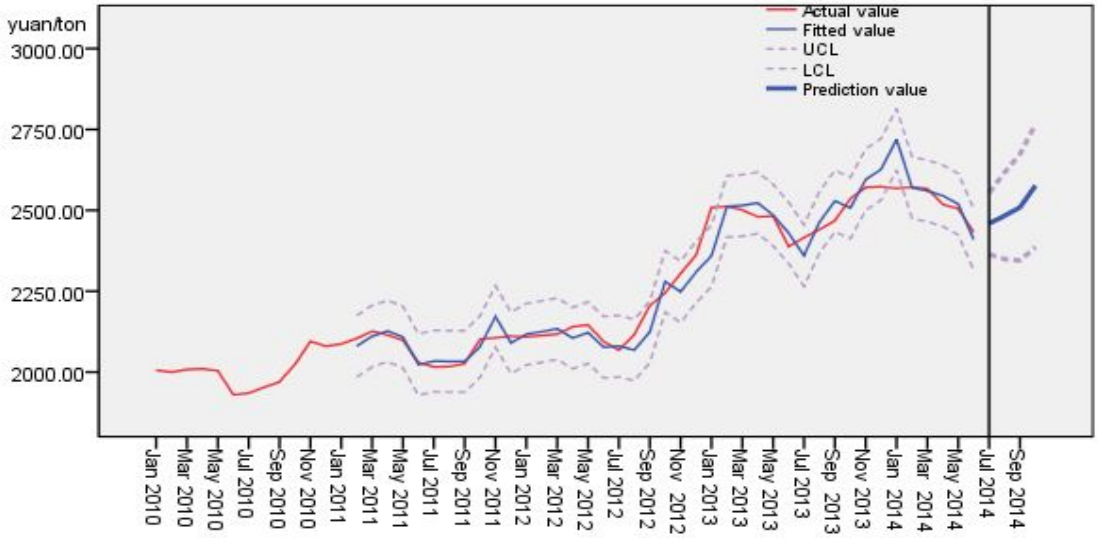



Fig.8 ARIMA (1,2,1) model predictions and fitting image

Table II ARIMA predictions results

| | **actual value** (yuan·ton$^{-1}$) | **prediction value** (yuan·ton$^{-1}$) | **Absolute error** | **relative error** |
|---|---|---|---|---|
| March | 2567 | 2520 | -47 | 1.83% |
| April | 2518 | 2502 | -16 | 0.64% |
| May | 2505 | 2492 | -13 | 0.51% |
| June | 2432 | 2426 | -6 | 0.25% |

## IV. CONCLUSION

Because use of ARIMA model to predict without considering other factors. Only the starting sequence itself, establish the appropriate model to predict. This at fundamental prevents the difficulty of identifying the main factors and secondary factors. For the time series in wheat prices and mainly showing the short-term correlation. therefore use time series to short-term prediction effect is better.

Finally, given four months to predict the future price

of wheat situation, we can found that in the next four months the price of wheat have growing trend. Above the predicted results can provide a basis for the development of Chinese government policies, adopt appropriate policies to ensure the stability of the market price of wheat, For ensuring food security and social stability is of great significance.

## ACKNOWLEDGMENT

This work supported by National High Technology Research and Development Program(2012AA10608).

## REFERENCES

[1] Yan Wang, Application of Time Series Analysis. China Renmin University Press Co. LTD. Beijing, 2010.
[2] Lina Shao, "Based on ARMA model to predict the stock price of China Merchants Bank," Rural Economy and Science-Technology. Hubei. pp.105-107, December 2007.
[3] Lu Hou, "The ARIMA model based on short-term analysis forecasting of oil prices," Jinan University. Guangdong. June 2009.
[4] Danhui Yi, Data analysis and application of Eviews. China Renmin University Press Co. LTD. Beijing, 2008.
[5] Lijie Zhang, "Cotton price study and forecasting," Tientsin University. Tianjin. June 2012
[6] Shiying Zhang, Cointegration Theory and Volatility Models. Tsinghua University Press. Beijing.pp.48-88, 2009.
[7] Hongyu Pan,Time series analysis. University of International Business and Economics Press. Beijing. Pp.221-247, 2006
[8] Fangda Gu, Chengliang Fang, "Characteristics of Chinese agricultural price volatility analysis," Chinese Rural Economy, Beijing, pp.67-76, June 2010.
[9] Saghaian, S.; Reed, M.; Merchant,M.et al: Merchant Monetary Impacts and Overshooting of Agricultural Prices in an Open Economy, American Journal of Agricultural Economics, 2: 90-103, 2002.
[10] Araceli Bernabe, Esteban Martina, etc. "A mufti-model approach for describing crude oil price dynamics," physica vol. A 338,. pp. 567-584, 2004.

**Shiliang Du ,**male**,** Master**,** was born in Xiangxiang, Henan province, China in 1989. Research interest: Pattern Recognition, Machine Learning.

**Dexian Zhang** (1961 — ) ,male,Doctor,Professor, major in Computer intelligence technology, Artificial intelligence, pattern recognition.

# Fuzzy Proximal Support Vector Machine for the Imbalanced and Balanced Datasets

Chuandong Qin
School of mathematics and information science, North National University, Yinchuan 750021, P.R. China
Email:qinchuandong123@163.com

***Abstract*—Introducing two fuzzy membership functions for the different features datasets and gaining the fuzzy membership, we propose the Fuzzy Proximal Support Vector Machine (FPSVM) for the imbalance datasets. This kind of FPSVM based on the PSVM can deal with the overfitting phenomenon among the imbalance classification problem. specially, reducing the influence of the outliers and noise in imbalanced datasets. Different fuzzy memberships may have different effects on the separating hyperplane. Points close to the proximal hyperplane have more effects than those away from the plane. Experiment results on six benchmark datasets show the classification accuracy of this algorithm is improved. Compared with other Fuzzy Support Vector Machines (FSVM) and PSVM, each method has its own preponderance and shortcomings for the different datasets, but this algorithm have the more selection for the classification of the imbalanced datasets.**

***IndexTerms*—proximal support vector machine, fuzzy membership, fuzzy proximal support vector machine, Linear decaying, exponential decaying.**

## I. INTRODUCTION

Your Support Vector Machine (SVM) [1-3] is a widely used machine learning technique, which has been applied to many real-world classification problems in various domains. Based on the theory of SVMs, many scholars did the further research in data mining fields. J.A.K. Suykens (1999) proposed Least Squares Support Vector Machine (LSSVM) classifiers with an equation instead of the inequality of Lagrange Support Vector Machine (LSVM) [4]. After two years, Fung and Mangasarian (2001) proposed the Proximal Support Vector Machine (PSVM) [5], which may be thought of a kind of regularized LSSVM. PSVM requires the solution of a single set of linear equations and which can be considerably faster than the traditional SVMs. The solution of this linear system requires the dimension of the inversion of a matrix is smaller than in the case of SVMs, which yields a computational advantage. And Fuzzy Support Vector Machine (FSVM) is a variant of the SVM learning algorithm, which was originally proposed in [6]. In order to handle outliers and noises in imbalance datasets, FSVM technique assigns different fuzzy membership values (weights) for different training examples to reflect their importance and then incorporates these membership values into SVM learning algorithm to reduce the effect of outliers and noise. There are many ways to define the membership of a training example [7]-[8].

In this paper, according to the features of the imbalance dataset, we define a suitable membership function and proposed the FPSVM based on the PSVM algorithm. The memberships are closely related to the features of the dataset and reflect the contribution rate of any example point for the optimal proximal hyperplane. The proposed method enhances the PSVM algorithm in reducing the effect of outliers and noises in dataset. Experimental results show the FPSVM is superior to the PSVM and gains the more selection for the different imbalance datasets.

## II. FUZZY MEMBERSHIPS

When training the SVM with all kinds of datasets, you will find the different example point give the different contribution rates to the hyperplane of classifier. So does the PSVM, the examples closer to the proximal hyperplane are treated as more informative and assigned higher membership values, while the examples far away from the proximal hyperplanes are treated as less informative and assigned lower membership values. To the PSVM, the proximal hyperplanes cross the center of the cluster. The distance of each point to the proximal hyperplane and the optimization hyperplane can be expressed respectively as follows:

$$d_i^p = \frac{\omega \cdot \varphi(x_i) + b \pm 1}{\sqrt{\|\omega\|}} \quad (1)$$

$$d_i^0 = \frac{\omega \cdot \varphi(x_i) + b}{\sqrt{\|\omega\|}} \quad (2)$$

The distance of the proximal hyperplane to the optimization hyperplane is $d = \frac{1}{\sqrt{\|\omega\|}}$, where $\|\omega\|$ is the 2-norm, $d_i^p$ is the distance to the proximal hyperplane and $d_i^0$ is the distance to the optimization hyperplane. $(\omega \cdot b)$ can be gained from (10)-(11) and $\omega \cdot \varphi(x_i)$ can also be worked out with the kernel function: $K(A, A^T) = \varphi(A) \cdot \varphi(A^T)$

For the PSVM, we only need to compute the membership of the point between the two proximal hyperplane. So the membership should be subject to the following conditions:

$$s_i = \begin{cases} s_i, if\ d_i^0 \le d \\ 0, if\ d_i^0 > d \end{cases} \quad (3)$$

Under the above conditions, we only need to compute nearly half of the points in two cases: Linear-decaying membership function and exponential-decaying membership function

$$s_i = 1 - \frac{d_i^p}{\max(d_i^p + \Delta)} \quad (4)$$

$$s_i = 1 - \frac{2}{1+\exp(\beta d_i^p)}, \beta \in [0,1] \quad (5)$$

Matrix $S = diag(s_1, s_2, \cdots s_m)$ is a diagonal matrix about the dataset with nearly half of the zeroes. Obviously, the small positive value $\Delta$ is used to avoid the case where becomes zero and the constant $\beta$ determines the steepness of the decay. Some outliers or noise which is far away from the proximal hyperplane will be given a very smaller value.

## III. FUZZY PROXIMAL SUPPORT VECTOR MACHINE

Given the dataset:

$T = \{(x_{1M}, y_1), (x_2, y_2), \cdots (x_m, y_m)\}, x_i \in R^m, y_i \in \{-1, 1\}$, $i = 1, 2, \cdots, n$ .Based on the LSSVM classifiers of [4], the PSVM of Glenn Fung was aimed to build a decision function in the feature space $F : x_i \to \varphi(x_i)$ which can be express as follows:

$$\begin{cases} \min\limits_{(\omega,b,\xi)\in R^{n+1+m}} \frac{1}{2}(\omega^T\omega + b^2) + \frac{C}{2}\|\xi\|^2 \\ subject\ to\ D(\varphi(A)\omega + eb) = e - \xi \end{cases} \quad (6)$$

Where A is the dataset and $\xi$ is the slack variable. For simplicity, let $\varphi(A)$ be the map in a feature space, and $K = K(A, A^T) = \varphi(A) \cdot \varphi(A^T)$ .Figure 2 describe its geometric explanation in $R^2$ , the planes $\omega \cdot \varphi(x_i) + b = \pm 1$ around the points " " and "Δ" are pushed apart.

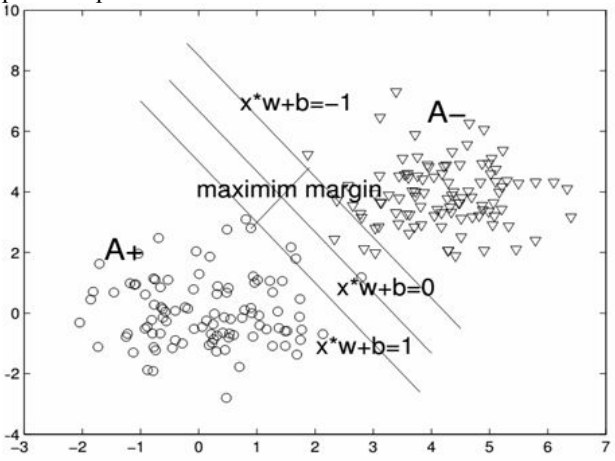



Figure1. FSVM Model

## IV. THE LINEAR FUZZY PROXIMAL SUPPORT VECTOR MACHINE

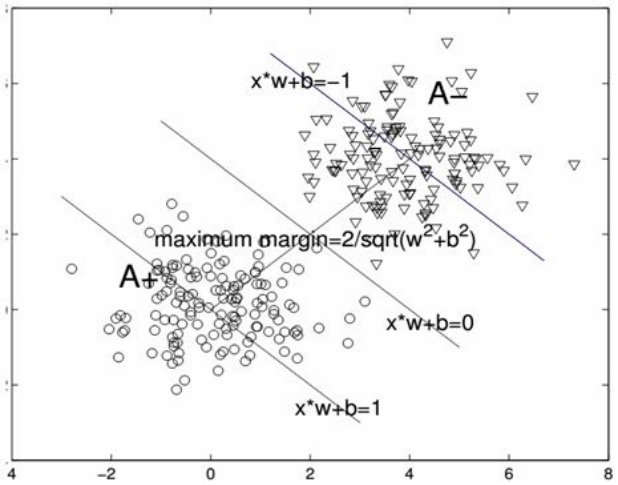



Figure2. PSVM Model

After gained the fuzzy membership, we propose the FPSVM as follows:

$$\begin{cases} \min\limits_{(\omega,b,\xi)\in R^{n+1+m}} \frac{1}{2}(\omega^T\omega + b^2) + \frac{C}{2}\|S\xi\|^2 \\ subject\ to\ D(A\omega + eb) = e - \xi \end{cases} \quad (7)$$

Here $S = diag(s_1, s_2, \cdots s_m)$ is the diagonal matrix. There are almost half of zeroes in the diagonal matrix *S* The diagonal elements correspond to its membership values, which are associated with a lower bound $\sigma \le s \le 1$ , where $\sigma > 0$ denotes the lower bound for the ith sample. To solve the model (7), we construct the following Lagrange function.

$$L(\omega, b, \xi, u) = \frac{1}{2}(\omega^T\omega + b^2) + \frac{C}{2}\|S\xi\|^2 - u^T D(A\omega + eb) - e + \xi. \quad (8)$$

Using KKT optimality conditions for (9), we can get the explicit formulation:

$$u = [HH^T + S^{-2}C^{-1}]^{-1}e \quad (9)$$

By letting $H = D[A - e]$ , we implement the Sherman-Morrison-Woodbury formula [12] to (8) and obtain

$$u = CS^2 - [CS^2H(I + H^TS^2H)^{-1}H^TCS^2]e \quad (10)$$

This expression includes an inversion of $(n+1)\times(n+1)$ matrix, which should be much simpler than the inversion of $m \times m$ matrix in the case of $n << m$ .The way computing (11) is the same to the linear PSVM. So the decision function for the linear fuzzy PSVM can also gains the same decision function: $f(x) = \text{sgn}(K(x^T, A^T)Du + b)$ .

## V THE NONLINEAR FUZZY PROXIMAL SUPPORT VECTOR MACHINE

Similar to the nonlinear PSVM algorithm, FPSVM algorithm can also be used in the nonlinear case. The equality constraint of (8) can be replaced by the following equality:

$$D(\varphi(A)^T\omega + eb) = e - \xi . \quad (11)$$

The nonlinear optimization problem for FPSVM is represented as follows:

$$\begin{cases} \min\limits_{(\omega,b,\xi)\in R^{n+1+m}} \dfrac{1}{2}(\omega^T\omega+b^2)+\dfrac{C}{2}\|S\xi\|^2 \\ subject\ to\ D(\varphi(A)^T\omega+eb)=e-\xi \end{cases} \quad (12)$$

Using KKT optimality conditions, we obtain an explicit expression

$$u=[GG^T+S^{-2}C^{-1}]^{-1}Se \quad (13)$$

Where $G$ is defined as $G=D[K,-e]$ .The nonlinear separating hyperplane corresponding to $\omega=\varphi(A)Du$ can be express $x^T\varphi(A)Du+b=0$ .Replacing $x^T\varphi(A)$ by the kernel function $K(x^T,A^T)$ we obtain the separating surface:

$$K(x^T,A^T)Du+b=(K(x^T,A^T)K(A^T,A^T)+e^T)Du=0 \quad (14)$$

The corresponding nonlinear classifier is

$f(x)=\mathrm{sgn}(K(x^T,A^T)Du+b)$.

Note that the above matrix $G$ and $H$ is similar. Unlike the situation with linear kernels,the SMW (Shermarr-Morrisorr-Woodbury) formula is useless, because the kernel matrix is a square matrix. The inversion in (13) cannot be converted to an inversion of $n\times n$ matrix as the linear case. So the solution of the nonlinear fuzzy PSVM need to apply the way of the nonlinear PSVM which adopt the reduce kernel technology. The computational complexity of the nonlinear FPSVM is the same as PSVM.

## VI. ALGORITHMS FOR FPSVM

*A. Training algorithm for linear Proximal SVM*

(i) Define $H$, compute $u$ by(13) for some positive C.and gain $(\omega,b)$ from PSVM.

(ii)Compute the distance of each point to the optimization hyperplane from(1)or(2)and gain the membership $S=diag(s_1,s_2,\cdots s_m)$ from(3)-(5).

(iii) Recomputed $u$ from (10) using the upper $S$, $H$ and u.

(iv) Determine $(\omega,b)$ from FPSVM and gain the linear classifier.

(v) Classify a new $x_i$ by using the linear classifier.

*B. Training algorithms for Nonlinear Proximal SVM*

We generate the nonlinear classifier as follows:

(i) Choose a kernel function $K(A^T,\overline{A}^T)$ , where $\overline{A}^T{}_{\overline{m}\times m}$ is the submatrix of $A$ and $\overline{m}$ is as small as 1% of m, typically the Gaussian kernel.

(ii) Define $G=D[K,-e]$ where $K=K(A^T,\overline{A}^T)$ and e is an $m\times 1$ vector. Compute $u$ by (13).

(iii) Compute the distance of each point to the hyperplane by (1) or (2) and gain the Membership $S=diag(s_1,s_2,\cdots s_m)$ by (3)-(5).

(iv) Recomputed $u$ by (13) using the upper $S$,$H$ and $u$.

(v) Determine $(\omega,b)$ from FPSVM and gain the nonlinear classifier.

(vi) Classify a new $x_i$ by using the nonlinear classifier.

## VII. EXPERIMENTS AND COMPARISONS

We consider five benchmark real-world datasets from the UCI machine learning repository [13] to validate the proposed FPSVM method. In the test, each dataset is scaled into [-1, +1] interval. and a five-fold cross-validation method is used. The results of the LSVM, PSVM and SSVM algorithm can be gain from [14] and [15].FSVMs is gained from [8] and [13].all of the results will be compared with the FPSVM in the following table.

Table I: Comparison of training accuracy, testing accuracy and running time (s) about PSVM, SSVM, LSVM and FPSVM using a linear kernel $k(x_i,x_j)=(x_i,x_j)$

| Dataset (m×n)P/N | PSVM | SSVM | LSVM | FPSVM | |
|---|---|---|---|---|---|
| | | | | Liear | Exp |
| WPBC (198×32) 151/47 | 70.8% | 70.8% | 70.8% | 72.5% | 71.2% |
| | 68.0% | 68.5% | 68.5% | 71.8% | 69.6% |
| | 0.02 | 0.17 | 0.53 | 0.30 | 0.48 |
| Ionosphere (351×34) 225/126 | 90.0% | 94.3% | 94.4% | 95.6% | 94.2% |
| | 87.0% | 88.7% | 88.7% | 94.2% | 93.4% |
| | 0.17 | 1.23 | 1.40 | 0.42 | 0.56 |
| WDBC (569×32) 357/212 | 77.0% | 78.2% | 78.2% | 79.5% | 79.2% |
| | 77.0% | 77.6% | 77.6% | 79.0% | 78.6% |
| | 0.02 | 0.78 | 2.18 | 1.06 | 1.56 |
| BUPA Liver (345×6) 145/200 | 70.8% | 70.8% | 70.1% | 71.3% | 71.2% |
| | 70.0% | 70.0% | 69.6% | 70.8% | 70.8% |
| | 1.75 | 1.05 | 0.34 | 1.98 | 2.40 |
| Mushroom(8124×22)4208/3916 | 81.0% | 81.7% | 81.7% | 84.6% | 82.8% |
| | 81.0% | 81.5% | 81.5% | 83.7% | 81.6% |
| | 1015 | 11.73 | 61.62 | 2836 | 3617 |

Table 1 shows the FPSVMs algorithm have the stronger processing ability. They gain the more precise results but the execution time increase much more than other SVMs,Because datasets have the different features.The linear decaying membership function is fit for the Ionosphere and Mushroom and the Exponential-decaying membership function is fit for the BUPA Liver and WPBC.Both of the two membership functions are effective for classifier.

Table II: Comparison of training accuracy, testing accuracy and running time (s) about LSVM, SSVM, PSVM and FPSVM using Gaussian kernel.

| Dataset (m×n)P/N | LSVM | SSVM | PSVM | FSVM | FPSVM | |
|---|---|---|---|---|---|---|
| | | | | | Line | Exp |
| Ionosphere (351×34) 225/126 | 97.0% | 97.0% | 96.5% | 93.5% | 96.7% | 97.8% |
| | 95.8% | 95.8% | 95.2% | 93.4% | 96.1% | 97.0% |
| | 14.57 | 25.25 | 4.60 | 2.89 | 6.80 | 7.68 |

| | | | | | | |
|---|---|---|---|---|---|---|
| BUPA Liver (345×6) 145/200 | 75.8% | 75.8% | 75.7% | 78.9% | 76.3% | 76.3% |
| | 73.7% | 73.7% | 73.6% | 78.8% | 75.9% | 76.1% |
| | 20.89 | 30.80 | 4.80 | 25.98 | 5.64 | 6.84 |
| Tic-Tac (956×9) 625/333 | 98.2% | 98.0% | 98.0% | 97.8% | 97.6% | 98.8% |
| | 94.7% | 98.4% | 98.4% | 97.0% | 97.4% | 98.0% |
| | 350.64 | 395.30 | 74.95 | 86.30 | 188 | 216 |
| Mushroon (812×22) 4208/3916 | 87.6% | 89.0% | 88.0% | 87.6% | 89.2% | 88.2% |
| | 87.8% | 88.8% | 88.0% | 87.2% | 88.7% | 87.4% |
| | 503.74 | 307.66 | 35.50 | 214 | 89.6 | 146 |

Table 2 tells us: The correctness of the five methods is very similar but the execution time of the FPSVM is larger than that of the PSVM and smaller than the other SVMS. it also shows the FPSVM is better than FSVM in some dataset, such as WPBC, Ionosphere, except for BUPA Liver. Those show the fuzzy membership function has the selectivity for different datasets. Some datasets are suit for our FPSVM algorithm; some datasets are suit for the FSVM.But most of FSVM can gain the better effect for the datasets with some noise or outliers especially.

## VIII. CONCLUSIONS

In this paper, we propose a FPSVM method with two types of fuzzy membership function for different datasets based on the PSVM algorithm. In this method, we assign a fuzzy-membership value for training examples according to the distance to the proximal hyperplane of the PSVM and validate the proposed method with several real-world datasets in two ways. From the overall results obtained, we can conclude that the proposed FPSVM method could result in significantly better classification results than PSVM, LSSVM and SSVM. But it spends more time than PSVM. Compared with the FSVM algorithms, FPSVM algorithm also has obvious superiority in selecting the fuzzy membership function. Those show us the FPSVM with different fuzzy membership function has a strong ability to handle the datasets with noise or outliers. As future work, it would be interesting to investigate the effectiveness of using the FPSVM method in the imbalance-learning or the twin support vector machine learning.

ACKNOLEDEGEMENT

Supported by the Natural Science Ningxia Foundation(NZ13095)

## REFERENCES

[1] V. Vapnik: the Nature of Statistical Learning Theory. Springer-Verlag, Berlin, Germany, 1995.
[2] C.Cortes,V.Vapnik:support vector networks. Machine Learning, vol. 20, pp. 273-297, 1995.
[3] J.Showe-Taylor, N. Christianini: Support Vector Machines and Other Kernel-based Learning Methods. Cambridge Univ.Press, Cambridge, U.K, 2000.
[4] J.A.K.SUYKENS, J.VANDEWALLE：Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 9: 293-300, 1999.
[5] G. Fung, O.L. Mangasarian: Proximal support vector machine classi1ers. in: D. Lee, et al. (Eds.), Procee-dings of theKDD-2001: Knowledge Discovery and Data Mining, San Francisco,California, Association for Computing Machinery, New York, 2001,pp. 77-86.
[6] Chun-Fu Lin, Sheng-De Wang: Fuzzy support vector machines. Transactions on neural networks, vol.13 (2), 2002, pp. 464-471
[7] Che-Chang Hsu, Ming-Feng Han, Shih-Hsing Chang, Hung-Yuan Chung:Fuzzy support vector machines with the uncertainty of parameter C.Expert Systems with Applications (2008), doi:10.1016 .2008.08.032.
[8] Pei-Yi Hao: Fuzzy one-class support vector machines. Fuzzy Sets and Systems 159 (2008) 2317-2336.
[9] Chun-fu Lin, Sheng-de Wang: Training algorithms for fuzzy support vector machine with noisy data, Pattern Recognition Letters 25(2004) 1647-1656.
[10] Chun-fu Lin, Sheng-de Wang: Fuzzy Support Vector Machines, IEEE transactions on neural networks, vol. 13, no. 2, march 2002.
[11] G.H.Golub and C. F. Van Loan: Matrix Computations. The John Hopkins University Press, Baltimore, Maryland, 3rd edition, 1996.
[12] Y.-J. Lee and O. L.Mangasarian: RSVM: Reduced support vector machines. Technical Report 00-07, Data Mining Institute, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, July 2000. Proceedings of the First SIAM International Conference on Data Mining, Chicago, April 5-7, 2001, CD-ROM.
[13] The UCI repository of machine learning database http://archive.ics.uci.edu/ml
[14] J.A.K. SUYKENS, J. VANDEWALLE: Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9: 293-300, 1999.Kluwer Academic Publishers. Printed in the Netherlands.
[15] Lee Yue Jye, O.L.Mangasarian: SSVM:A Smooth Support Vector Machine for Classification. Computational Optimization and Applications, 2001, 20:5-22.

# Journal of Simulation
## (JSM, ISSN: 2310-4791)
## CALL FOR PAPERS

**Journal of Simulation (JSM, ISSN: 2310-4791)** is a scholarly peer-reviewed international scientific journal published bimonthly for researchers, developers, technical managers, and educators in the simulation field. It provide a high profile, leading edge forum for academic researchers, industrial professionals,engineers, consultants, managers, educators and policy makers working in the field to contribute and disseminate innovative new work on all the areas of simulation.

*JSM* invites original, previously unpublished, research, survey and tutorial papers, plus case studies and short research notes, on both applied and theoretical aspects of simulation.

### Area of Interests

- Simulation
- Economics Simulation
- Management Simulation
- Model
- Operating Systems
- Software Systems
- Communication Protocols
- Interconnection Networks
- Design
- Prototyping
- Testing Methods and Tools
- Artificial Intelligence
- Algorithms
- Computational Science
- New and Important Applications and Trends

### Manuscript Submission

- JSM invites original, previously unpublished, research papers, review, survey and tutorial papers, application papers, plus case studies, short research notes and letters, on both applied and theoretical aspects.
- Submission implies that the manuscript has not been published previously, and is not currently submitted for publication elsewhere. Submission also implies that the corresponding author has consent of all authors.
- Upon acceptance for publication transfer of copyright will be made to Academic Publishing House, author will be requested to pay a mandatory publication charge. Article submission implies author agreement with this policy.
- Whenever applicable, submissions must include the following elements: title, authors, affiliations, contacts, abstract, index terms, introduction, main text, conclusions, appendixes, acknowledgement, references, and biographies.

### Publication Charges and Orders

All the journals published by Academic Publishing House are 'Open Access' journals, which means that the publication cost should be covered by the author's institution or research funds. The author's company or institution will be requested to pay a flat publication fee of EUR 200 for an accepted manuscript regardless of the length of the paper. The publication charges are mandatory. Authors can order hard copies of the issue with a price of EUR 20 per copy.